CONSTRUCCIÓN DE UNA MÉTODOLOGIA EMPLEANDO LA HERRAMIENTA R PARA ESTIMAR VALORES DE LOS ACTIVOS BANCOLOMBIA, BOGOTÁ Y OCCIDENTE CON MODELOS ARIMA

JENNYALEJANDRA ARDILA ARIZA

EDITH AVILA GOMEZ

ESTUDIANTES DE ESPECIALIZACION EN ESTADISTICA APLICADA

FUNDACION UNIVERSITARIA LOS LIBERTADORES
PROGRAMA DE ESTADISTICA APLICADA
BOGOTA
2017

CONSTRUCCIÓN DE UNA MÉTODOLOGIA EMPLEANDO LA HERRAMIENTA R PARA ESTIMAR VALORES DE LOS ACTIVOS BANCOLOMBIA, BOGOTÁ Y OCCIDENTE CON MODELOS ARIMA

PRESENTADO POR

JENNYALEJANDRA ARDILA ARIZA

EDITH AVILA GOMEZ

ESTUDIANTES DE ESPECIALIZACION EN ESTADISTICA APLICADA

ASESOR

MANUEL FRANCISCO ROMERO OSPINA

FUNDACION UNIVERSITARIA LOS LIBERTADORES

PROGRAMA DE ESTADISTICA APLICADA

BOGOTA

2017

Nota de Aceptación

Firma del presidente del jurado

Firma del Jurado

Firma del Jurado

Bogotá, D.C, 11de Diciembre de 2017

Las Directivas de la Universidad de
Los Libertadores, los jurados calificadores y el cuerpo
docente no son responsables por los
criterios e ideas expuestas en el presente documento.
Estos corresponden únicamente a los autores.

TABLA DE CONTENIDO

INDICE DE FIGURAS

INDICE DE TABLAS

CONSTRUCCIÓN DE UNA MÉTODOLOGIA EMPLEANDO LA HERRAMIENTA R PARA ESTIMAR VALORES DE LOS ACTIVOS BANCOLOMBIA, BOGOTÁ Y OCCIDENTE CON MODELOS ARIMA

** Jenny Alejandra Ardila Ariza - Edith Ávila Gómez

RESUMEN

En finanzas las acciones varían diariamente, bajando o subiendo su valor económico, pero saber si es rentable una acción en un periodo de tiempo puede llegar a ser un problema, uno de los métodos más utilizados es predecir el valor futuro de estas acciones para conocer si su valor económico crece o decrece, de esta manera se puede saber qué tan rentable es invertir. El propósito del presente trabajo es la construcción de una metodología empleando la herramienta R para estimar valores de los activos de las entidades Bancolombia, Bogotá y Occidente. Esta metodología ayudará por medio de las series de tiempo ARIMA, a pronosticar el precio del valor futuro de las acciones de los bancos encontrando un intervalo de máxima amplitud. Se evaluará la normalidad de los residuales de la series de las tres entidades, posteriormente se realizara un análisis detallado del intervalo de mayor amplitud que se ajuste a los datos reales con el menor error posible será escogido como el intervalo optimo a simular y su amplitud será escogida para simular otros intervalos y se verificará la gráfica de los datos simulados con los datos reales comparando el error observando la eficiencia de la metodología

.

PALABRAS CLAVES: Predicción con series de tiempo, valores de tendencia, programa R, predicción, ARIMA, activos.

ABSTRACT

Finance stocks vary daily, lowering or increasing their economic value, but to recognize if a stock is profitable in a period of time can be a problem, one of the most used methods is to predict the future value of these actions to understand if its economic value grows or decreases, in this way you can know how profitable it is to invest. The purpose of this investigation is the construction of a metodology that uses the software R in order to estimate values of the actives of the entities Bancolombia, Bogotá and Occidente. This methodology will help, by means of the ARIMA time series, to foresee the price of the future value of the actions of the banks, finding anmaxim amplitude interval. The normality of the residuals of the series of the three entities will be evaluated, afterwards a detailed analysis of the interval of greater amplitude that is adjusted to the real data will be realized with the smaller possible error which will be selected like the optimum interval to be simulated and its amplitude will be chosen to simulate other intervals and the graph of the simulated data with the actual data will be verified by comparing the error and checking the efficiency of the methodology.

KEYWORDS: Prediction with time series, trend values, Software R, prediction, ARIMA, actives.

**Estudiantes de especialización en Estadística Aplicada

1. INTRODUCCION

En finanzas las acciones varían en cada instante de tiempo, bajando su valor económico como también pueden llegar a subir de precio, pero saber si es rentable una acción puede llegar a ser un problema, uno de los métodos más utilizados es predecir el valor futuro de estas acciones para conocer si su valor económico crece o decrece, de esta manera se puede saber qué tan rentable es invertir.

Existen distintos métodos que hacen uso de los procesos estocásticos y nos ayudan a predecir el futuro de las acciones, pues en la economía al igual que en la naturaleza muchos fenómenos tienen un comportamiento de ruido blanco, entre esos modelos encontramos los modelos ARIMA (Autoregressive Integrated Moving Average) que con base en datos recopilados simula una serie temporal del activo y predice el valor numérico que puede llegar a tener esta serie en un futuro.

En Colombia se cuenta con una Bolsa de Valores en la que realizan negociaciones de compra y venta de acciones, bonos, títulos de deuda pública, entre otras que se puedan ofrecer dictaminadas por el gobierno, existían la bolsa de Medellín y la bolsa de Bogotá, entre otras.

Actualmente, todas se han unificado en una sola: la Bolsa de valores Colombia (BVC), la cual tiene su sede principal en Bogotá, en donde se cuentan con profesionales especializados en la valoración de activos por medio de modelos estocásticos.

En el presente trabajo se muestra una metodología que ayudará por medio de las series de tiempo ARIMA a valorizar activos en un tiempo futuro, esta metodología podrá ser implementada en cualquier tiempo, dado que se presentan los pasos a seguir para poder encontrar un intervalo de máxima amplitud que me permita predecir el valor futuro del activo en algún tiempo (determinado por la cantidad de datos).

La metodología se basa en la realización de simulaciones tomando el valor del precio de cierre en intervalos de tiempo de distinta longitud de tres bancos que servirán como ejemplo en la metodología, los cuales son Bancolombia, Bogotá y Occidente, dichos intervalos (tomados de los últimos 10 años).

Estos datos serán simulados en R para que éste nos arroje cuál es el modelo ARIMA que mejor se ajusta a los datos, luego se calculará el valor futuro con el intervalo simulado en el modelo ARIMA que se obtuvo, comparando los datos simulados con los reales.

Para el intervalo de mayor amplitud que se ajuste a los datos reales con el menor error posible será escogido como el intervalo optimo a simular y su amplitud será escogida para simular otros intervalos y se verificará la gráfica de los datos simulados con los datos reales comparando el error observando la eficiencia de la metodología.

Los modelos ARIMA en series de tiempo pueden ser utilizados para simular la media de los retornos financieros y para encontrar dicha parte predecible podemos hacer uso de programas estadísticos como lo es R mediante el paquete forecast, lo que facilita la programación y por ende la obtención de los resultados de predicción.

Aunque no siempre podemos utilizar los modelos ARIMA con todos los datos que deseemos, primero se debe verificar que los datos satisfagan las condiciones del modelo, posteriormente se debe simular en periodos ya conocidos para verificar que el error sea bajo con respecto a los datos reales. Pues se puede tener que los datos sean ajustados por una serie pero que se obtenga una mala predicción.

El modelo ARIMA identifica los coeficientes y número de regresiones que se utilizarán, este modelo es muy sensible a la precisión con que se determinen sus coeficientes.

El propósito definitivo del documento es presentar una la metodología que permita simular el valor futuro de una acción únicamente con modelos ARIMA, sin que esté presente mixturas ni variaciones al modelo original lo cual no es común pues existen otros métodos más complicados que permiten valorar activos.

Una de las limitaciones que presenta este trabajo es la posibilidad de obtener una predicción de amplia longitud, pues en la práctica consideramos que si se tienen n datos se podrá predecir el futuro hasta un tiempo máximo n, lo cual no sucede comúnmente, pues lo usual es que el tiempo de predicción no llegue a la misma cantidad de los datos. Para un intervalo de máxima amplitud de longitud muy pequeña la predicción será muy cercana al presente y no permitirá tomar una pronta decisión con respecto al activo, no siendo muy útil obtener dicha predicción.

2. PLANTEAMIENTO DEL PROBLEMA

Existen diversos métodos complejos para la determinación de un valor futuro de las acciones, aunque actualmente existen distintas herramientas computacionales que nos permiten obtener diversas predicciones, es necesario tener claro los procedimientos que estás realizan al ser utilizadas. Con este documento se pretende presentar una metodología para que cualquier persona con conocimientos en series de tiempo pueda obtener la predicción de un valor futuro mediante una serie ARIMA.

3. FORMULACION DEL PROBLEMA

¿Cuál es la predicción del precio de cierre de las acciones de la Bolsa de Valores de Colombia (BVC) de los bancos Bancolombia, Occidente y Bogotá entre el 2006 y el 2016, aplicando un modelo ARIMA de serie de tiempo que se ajuste adecuadamente, empleando la herramienta de R para determinar los intervalos de máxima amplitud que pronostiquen el valor futuro del precio de cierre de cada uno de los bancos?

4. JUSTIFICACIÓN

En la investigación se pretende predecir los valores de cierre de las acciones de los Bancos Bancolombia, Bogotá, y occidente se hará utilizando un modelo ARIMA ya que los precios de las acciones son series con tendencias estocásticas y esta es una característica fundamental de las series que siguen dichos modelos.

Además, en las series con tendencia estocástica existen períodos de duración aleatoria donde la serie crece linealmente y es algo que sucede muy a menudo en los precios de las acciones.

Aplicado a nuestro trabajo, el precio de las acciones de los bancos pude crecer linealmente por efecto de la demanda, pero puede llegar un punto donde la demanda alcanza su límite y los precios pueden empezar a oscilar entre ciertos valores o tener otro tipo de comportamiento.

Este fenómeno es conocido como acumulación de volatilidad, que se caracteriza porque existen intervalos de tiempo prolongados donde el precio de los activos presenta grandes variaciones en su precio seguido de periodos de tiempo de relativa tranquilidad.

Mencionada la importancia de los modelos ARIMA en el pronóstico económico y en especial en los precios de las acciones, se calculará con ayuda de la metodología de BJ y el lenguaje de programación R, el intervalo de máxima amplitud que pronostique el valor futuro del precio de cierre de cada uno de los bancos.

La escogencia del último enfoque para predecir los valores de cierre de las acciones de los bancos se debió básicamente a que los demás enfoques tienen ciertos problemas como por ejemplo: aumento del error de pronóstico si se va demasiado lejos hacia el futuro (modelos de regresión uniecuacionales), parámetros estimados no son invariantes ante cambios de política (modelos de regresión de ecuaciones simultáneas), selección del rezago que influirá en el número de parámetros a estimar y que podría ser un número grande (VAR).

Además, escogimos los modelos ARIMA o Metodología de Box- Jenkins (BJ) como también se conoce, como eje central en nuestro trabajo puesto que esta metodología no se deriva de teoría económica alguna ya que los valores de una variable Yt que se pronostican se explican a partir de valores pasados o rezagos de sí misma y por los términos de errores estocásticos.

También, permite transformar series no estacionarias (muchas series económicas son de este tipo) en estacionarias para suponer que sus características son constantes en el tiempo y en particular en el futuro donde se pretende pronosticar. Igualmente, la metodología es clara en cuanto a sus etapas de ejecución que una vez llevadas a cabo de forma iterativa arrojen un modelo ARIMA que sirva para pronosticar.

La investigación que se presenta en el documento se realiza con el fin de obtener una metodología sencilla que permita simular el valor futuro de un activo sin necesidad de recurrir a modelos complejos o desconocidos para personas no especializadas en la valorización de activos.

Para hacer un pronóstico económico de una serie de tiempo como el que se pretende hacer en este trabajo, existen diferentes enfoques entre los que se destacan:

- Métodos de suavizamiento exponencial,
- Modelos de regresión uniecuacionales,

- Modelos de regresión de ecuaciones simultaneas,
- Modelos de vectores autorregresivos (VAR)
- Modelos autorregresivos integrados de promedios móviles (ARIMA).

Esta metodología es presentada por primera vez en este trabajo pues actualmente herramientas computacionales como R facilitan la obtención de los intervalos que son la parte principal de la investigación, beneficiando a investigadores que no tengan conocimientos en modelos basados en procesos y simulación estocástica (diferente al ARIMA) permitiéndoles replicar la metodología para hallar la predicción no solo de activos sino que también de otros datos que se ajusten a los modelos.

5. OBJETIVOS

5.1.OBJETIVO GENERAL

- Proponer una metodología para hallar modelos ARIMA de serie de tiempo que se ajuste a los valores de las acciones de la bolsa de Valores de Colombia de los bancos de Bancolombia, Occidente y Bogotá, que permita pronosticar el valor futuro del precio de cierre de cada uno de los bancos.

5.2.OBJETIVOS ESPECÍFICOS

- Plantear un modelo ARIMA de ajuste mediante el uso de la herramienta de R, que simule los intervalos de máxima amplitud, simulando el valor futuro de los bancos de Bancolombia, Occidente y Bogotá con una cota de error.
- Analizar el modelo ARIMA de ajuste de series de tiempo para predecir el comportamiento de los valores futuros de los bancos de Bancolombia, Occidente y Bogotá.
- Proponer un modelo ARIMA de series de tiempo que se ajuste a los intervalos de máxima amplitud que nos permita predecir el valor futuro de las acciones de los bancos de Bancolombia, Occidente y Bogotá.

6. MARCO DE REFERENCIA

En el presente trabajo que pretende ser autocontenido, se presentan a continuación algunas definiciones tomadas de algunos textos sobre series de tiempo, con el fin de que el lector familiarizado o no en el tema, las tenga al alcance siempre que lo considere necesario para el entendimiento de lo se está haciendo, sin necesidad de recurrir a bibliografía externa.

A continuación, se mostrarán los diferentes conceptos utilizados en este proyecto, en primera instancia se definirá al programa R para series de tiempo, en segunda instancia la naturaleza estocástica y sus componentes, en tercera instancia se definirá qué es una serie de tiempo y cuáles son sus componentes y en cuarta instancia se precisará sobre los métodos estadísticos propuestos para determinar cuál es el intervalo de máxima amplitud que permita pronosticar.

6.1.PRELIMINARES

Dada la naturaleza estocástica está involucrada al tratar de calcular precios de acciones tal y como se pretende en este y otros escritos, es más que necesario ver qué características tienen los procesos que describen este comportamiento y que nos permiten hacer predicciones o pronósticos probabilísticos. En nuestro caso, los procesos estocásticos que consideramos son los precios de cierre de las acciones de los bancos (BBO), que van a ser vistos como series de tiempo debido a que es lo usual en la literatura y además nos permite hacer interpretaciones precisas sobre los resultados obtenidos. Con esto en mente se introducirán ciertos conceptos que van a ser explicados a partir de la utilidad en nuestro trabajo.

6.2.PRELIMINARES COMPUTACIONALES [16]

La simulación en computación ha sido de gran ayuda al momento de realizar procesos numéricos de gran complejidad, sin embargo, esta cuenta con errores intrínsecos al momento de arrojar los resultados, estos errores son a pequeña escala y dependen de la aproximación que la máquina tome al momento en que se introduce un número. Puesto que las máquinas tienen una memoria limitada, también lo es el número de dígitos que estas pueden procesar. Si bien sabemos $\frac{1}{3} = 0,\hat{3}$, es un decimal periódico con términos infinitos, una máquina representa esa fracción como un decimal finito no periódico donde el último número de esta expansión ya no es 3. Siendo así, la acumulación de operaciones con números aproximados genera un error en el resultado, este depende de cuales sean los números con los que se opera. A continuación, se presentarán algunos conceptos que ayudarán a esclarecer el concepto de los errores computacionales generados por la simulación.

6.2.1. Punto Flotante

Forma de notación científica usada en los microprocesadores con la cual se pueden representar números racionales extremadamente grandes y pequeños de una manera eficiente y compacta.

Los números de punto flotante son la representación del número en la máquina, y estos me generan algunos errores representados como se sigue,

- (Error de redondeo.) Este error es debido a la aproximación que genera la máquina en la representación del número como punto flotante.
- (Error de propagación.) Es el error que se genera cuando la máquina realiza operaciones entre números aproximados por la máquina, incrementado el error en cada procedimiento por el error de redondeo.
- (Error de comparación.) Por los errores que se generan en el error de redondeo al realizar operaciones que nos den un resultado aproximado al resultado real, y al comparar estos dos valores no se obtendrá que sean iguales (pero si próximos).

Con esta sección se busca que el lector comprenda mejor porque se generan errores en la simulación y por qué al momento de simular se buscan buenas aproximaciones de los datos buscados, mas no valores exactos. A continuación, se presenta una breve introducción al programa R, que es un entorno y lenguaje de programación con enfoque estadístico.

6.3.INTRODUCCIÓN AL PROGRAMA R PARA SERIES DE TIEMPO.

R es una implementación de software libre de lenguaje S (statical), que es un lenguaje de programación pensado para un fácil procesamiento de los datos y plasmar muchas ideas de una forma más sencilla a diferencia de lenguajes basados en C o Java que son un poco más complejos. El programa R representa una gran ventaja en cuanto nos referimos a software libre tanto en el programa como en sus librerías, pues es posible trabajar con otros programas que permiten procesar una mayor cantidad de datos, tales como Matlab o Mathematica y tienen toolbox en estadística, sin embargo, estos no son versiones gratuitas y pueden llegar a tener costos muy elevados. Por esta razón en esta sección se presenta una breve introducción a R, haciendo uso de una de sus librerías para el análisis que se pretende realizar.

Para el análisis de las series de tiempo con el programa R, existe una librería llamada **forecast** (Forecasting Functions for Time Series and Linear Models), esta librería nos provee métodos y herramientas para mostrar y analizar pronósticos de series de tiempo univariadas incluyendo el suavizado exponencial a través de modelos de espacio de estado y modelado ARIMA automático.

De esta librería se usarán algunos comandos que serán explicados con base en el documento de la librería forecast [15] proporcionado por R.

6.3.1. Accuracy

Devuelve el rango de medidas de resumen de la precisión de pronóstico. Si se proporciona x, la función mide fuera de la muestra (conjunto de pruebas) precisión de pronóstico basada en x-f. Si x no se proporciona, la función sólo produce en la muestra (conjunto de entrenamiento) las medidas de precisión de los pronósticos basados en f["x"] - fitted(f) (función genérica que extrae los valores ajustados de los objetos devueltos por las funciones de modelado.)

Devuelve las medidas:

- ME: Error promedio.
- RMSE: Error cuadrático medio.
- MAE: Error medio absoluto.
- MPE: Error medio porcentual.
- MAPE: Error medio absoluto porcentual.
- MASE: Error medio absoluto escalado
- ACF1: Autocorrelación de errores al retraso.

6.3.2. Auto.arima

Devuelve el mejor modelo ARIMA según el valor AIC, AICc o BIC. La función lleva a cabo una búsqueda sobre posible modelo dentro de las restricciones de orden proporcionadas, donde:

- AIC: Criterio de información de Akaike $\left(2k - 2\ln(L)\right)$ que proporciona un medio para la selección de un modelo basándose en los k parámetros del modelo estadístico y en el máximo valor L de la función de verosimilitud para el modelo estimado. El modelo con menor valor AIC es el modelo escogido como optimo.

- AICc: Es una corrección de AIC para pequeñas muestras (de tamaño n) dado por;

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}. \qquad (1)$$

- BIC: Criterio de información bayesiano $(-2\ln(L) + k\ln(n))$, es una consecuencia derivada asintótica bajo los supuestos de que la distribución de los datos se encuentra en la familia exponencial, este modelo usa los mismos parámetros de entrada que AICc.

6.3.3. Forecast

Es una función genérica para la predicción de series de tiempo o modelos de series de tiempo. La función invoca métodos particulares que dependen de la clase del primer argumento, es decir realiza la predicción basándose en los datos obtenidos por el comando auto.arima aplicado en la serie.

Con estos comandos introducidos se pretende obtener resultados de predicción para las series de los bancos mencionados anteriormente, y así haciendo uso de los avances computacionales, encontrar el intervalo de mínima amplitud que me genere la mejor predicción, viendo esta con una comparación de los errores entre el valor real y el valor obtenido en la predicción.

6.4.DEFINICIONES.

6.4.1. Espacio de Probabilidad

A la tripla (Ω, F, P) se le conoce como espacio de probabilidad conformado por el espacio muestral Ω qué es donde están todas las variables a considerar, la sigma álgebra

F que es un conjunto de eventos (variables) que pueden tener cero o más resultados y P una medida de probabilidad, que asigna probabilidades a los eventos. En el espacio de los precios de los activos estarán presentes las series a considerar.

6.4.2. Proceso Estocástico

Un proceso estocástico o aleatorio1 (Damodar N. Gijarati, Dawn C. Porter, 2009), es una colección de variables aleatorias ordenadas en el tiempo. Los procesos pueden ser discretos o continuos de acuerdo a si las variables en consideración son discretas o continuas respectivamente. Estas variables suelen tener subíndices de numeración que hacen referencia a una observación específica en el tiempo. Cuando la variable toma un valor numérico en concreto, se hablará de una realización de la variable.

6.4.3. Incrementos Independientes e Incrementos Estacionarios

En un proceso en el cual los cambios en intervalos que no se solapan, la probabilidad conjunta de estos intervalos se puede expresar como el producto de las probabilidades de los intervalos, se dice que los incrementos son independientes. Formalmente2 (Liliana Blanco Castaneda V. A., 2012), si para todo $t \in A \subset \Re^{+}$, la variable aleatoria $X_{t+s} - X_s$ es independiente de X_s para todo $s < t$, entonces el proceso estocástico $\{X_t ; t \in A\}$ se dice que es un proceso con incrementos independientes.

Un proceso estocástico se dice que es estacionario si cambios en el proceso de igual tamaño son iguales en distribución. Debidamente, un proceso estocástico $\{X_t ; t \in A\}$ se dice de incrementos estacionarios si $X_{t_{i+1}+r} - X_{t_i+r}$ tiene la misma distribución que $X_{t_{i+1}} - X_{t_i}$ para todos los tiempos t_i y $r > 0$.

Ahora se procede a introducir la definición de *Movimiento Browniano* (MB), fenómeno observado por Robert Brown al notar el movimiento incesante de las partículas de polen suspendidas en un líquido pero que no fue sino hasta 1905 que Albert Einstein desarrolló toda la teoría matemática para explicarlo3 (Liliana Blanco Castaneda V. A., 2012).

Ahora sabemos que se presenta en la naturaleza de forma regular al igual que en situaciones financieras, matemáticas y muchas otras, haciendo de este un proceso de gran interés en la actualidad.

6.4.4. Movimiento Browniano

(Liliana Blanco Castaneda V. A., 2012). **4**El proceso $\{B_t : t \in [0,\infty)\}$ sobre (Ω, F, P), es un Movim2iento Browniano (MB) si:

1. $B_0 = 0$, es decir $P\{\{\omega : B_0(\omega) \neq 0\}\} = 0$.

2. B_t ~Nor$(0,t)$. Si $s < t$ se tiene que $B_t - B_s \sim$ Nor$(0, t-s)$.

3. $\{B_t\}_{t\geq 0}$ tiene incrementos independientes y estacionarios.

4. Para cada $\omega \in \Omega$ se tiene que $B_t(\omega)$ es una función continua.

En la anterior definición Nor denota la distribución Normal. En la Figura 1. Se ilustra una simulación del Movimiento Browniano, podemos observar que el MB empieza en 0, es una función continua en todo punto, es decir, su gráfica puede realizarse en un solo trazo. Tampoco es una función monótona en ningún intervalo, quiere decir que si tomamos un intervalo de cualquier longitud, por más pequeño que sea, la función no va a ser creciente ni decreciente; además tampoco va a ser diferenciable en ningún punto puesto que sin importar e zoom que hagamos a la gráfica esta nunca será suave y tendrá picos lo que no nos permitirá hallar la derivada en ningún lugar.

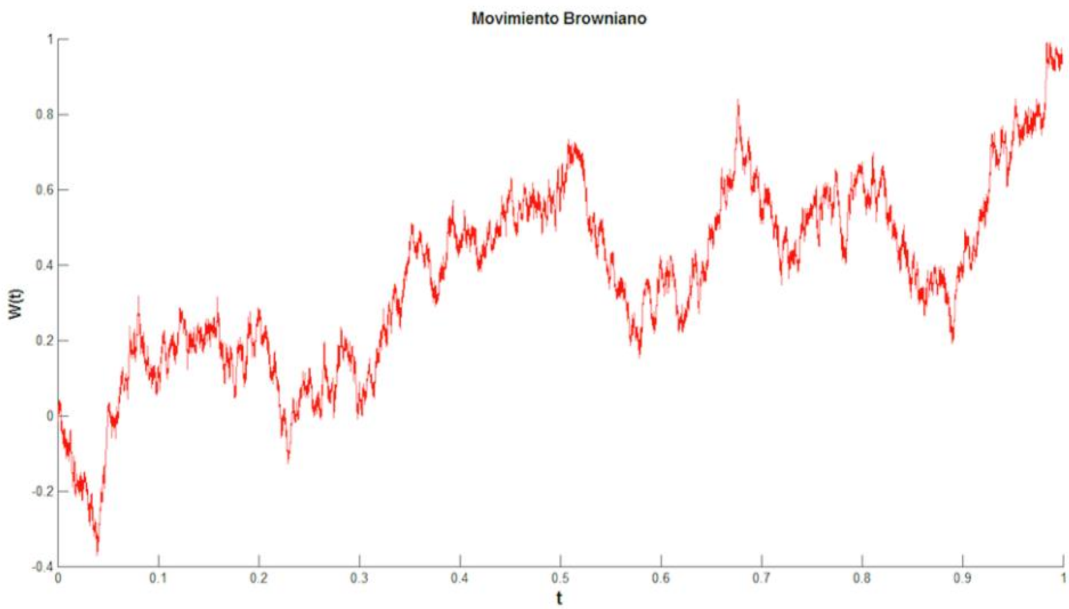

Figura 1. Simulación del MB con 2000 iteraciones en un intervalo de tiempo [0,1].

6.4.5. Ruido Blanco

Definimos el ruido blanco u_t como la derivada del MB en un sentido adecuado 5 (Todorovic, 2008), pues como ya se dijo, el MB no es diferenciable en ningún punto del tiempo. Es decir,

$$u_t = \frac{dB_t}{dt}$$

Alternativamente también diremos que una serie de tiempo (que será definida a continuación) es ruido blanco si es de media cero, varianza constante y no existencia de correlación para elementos diferentes de la serie. Formalmente lo anterior, para una serie $(\varepsilon_t, t \in Z)$ es: [6]

1. $E(\varepsilon_t) = 0$, media cero.

2. $Var(\varepsilon_t) = \sigma^2$, varianza constante.

3. $\forall k \neq 0, Cov = (\varepsilon_t, \varepsilon_{t+k}) = 0$ No correlación

Una vez introducidos los preliminares nos adentramos en nuestro concepto clave, las "Series de Tiempo", en especial las series de tiempo ARIMA.

6.5.SERIES DE TIEMPO

Una serie de tiempo es un conjunto de observaciones sobre los valores de una variable en diferentes momentos. Tal información debe recopilarse en intervalos regulares, es decir, en forma diaria (precios de acciones, informes del tiempo, etc.), semanal (como cifras de oferta monetaria), mensual (tasa de desempleo, Índice de Precios al Consumidor [IPC], etc.), trimestral (como el PIB), anual (como los presupuestos del gobierno), quinquenal (como el censo de la industria manufacturera), o decenal (como los censos de población). Algunas veces los datos están disponibles por trimestre y por año, como los datos del PIB y del consumo. Con las computadoras de alta velocidad, ahora se recopilan datos en intervalos muy breves, por ejemplo, precios de acciones, que se obtienen literalmente de manera continua (o cotización en tiempo real).

En la Figura 2. podemos observar un ejemplo de una serie de tiempo con base en los históricos de la Energía, aquí, el interés es implementar modelos de regresión sobre la serie histórica de precios de bolsa de energía en Colombia con el fin de explicar el comportamiento del mercado y ver los patrones que tienen las predicciones de los precios de la Energía en Bolsa.

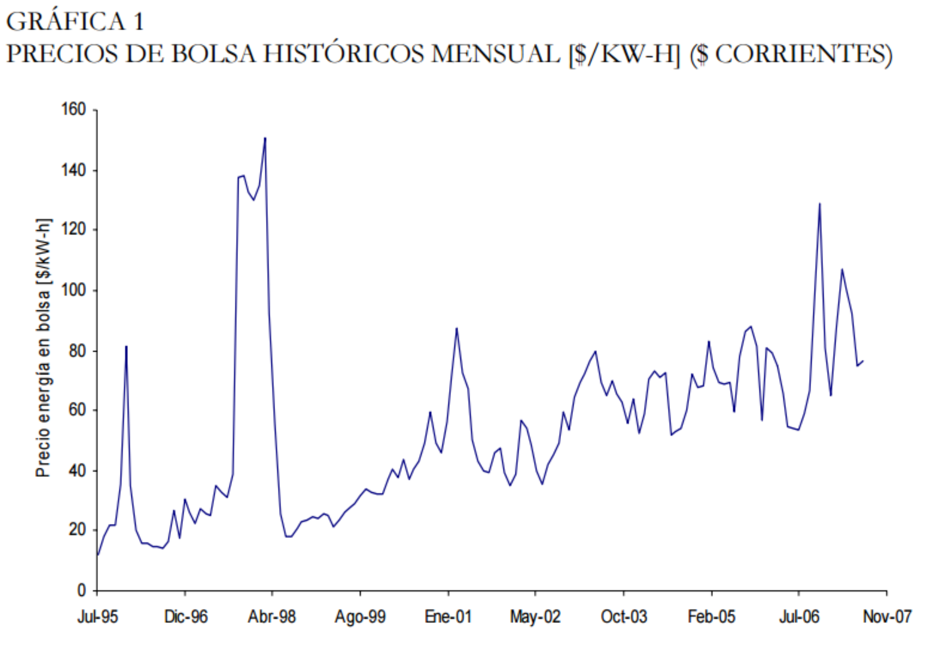

Figura 2. Serie de tiempo de los precios históricos de la Energía en la Bolsa Mensual en $/kW-h ($ por kilovatio - hora)

6.5.1. Serie Estacionaria

Una serie de tiempo es estacionaria si su media, su varianza y su autocovarianza (en los diferentes rezagos) permanecen iguales sin importar el momento en el cual se midan; es decir, son invariantes respecto del tiempo.

6.5.2. Objetivos específicos de las series de tiempo

Sánchez (2008) menciona que entre los objetivos específicos de las series de tiempo se encuentran:

- Obtener modelos estadísticos que describan la estructura pasada de las observaciones que generan la serie.
- Suponer que la estructura pasada de la serie de interés se conserva y bajo este supuesto, pronosticar valores futuros de la serie bajo estudio.
- Analizar la significancia de los efectos que causaron las intervenciones en la estructura de la serie.
- Simular valores futuros de la serie, bajo condiciones o restricciones definidas por criterios nuevos, para así supervisar y controlar los cambios que se producen en la serie.

6.5.3. Componentes de una serie de tiempo

- Tendencia: Es la componente de largo plazo que constituye la base del crecimiento o declinación de una serie histórica.
- Ciclicidad: Conjunto de fluctuaciones en forma de onda o ciclos, de más de un año de duración, producidos por cambios en las condiciones económicas.
- Estacionalidad: Las fluctuaciones estacionales se encuentran en los datos clasificados por periodos de tiempo. La variación estacional se refiere a un patrón de cambio, regularmente recurrente a través del tiempo.
- Aleatoriedad: Este comportamiento irregular está compuesto por fluctuaciones causadas por sucesos impredecibles o no periódicos, por ello las series de tiempo cuentan con una componente estocástica (ruido blanco).

Por estos comportamientos de la serie de tiempo, que no se evidencian solo con los datos es importante verificar la gráfica que producen estos datos, pues la importancia al analizar la gráfica de la serie según como se indica en [7], es que al hacer una gráfica de la serie se puede ver si esta tiene alguna tendencia que posiblemente sugeriría de forma intuitiva una variación de la media.

6.5.4. Función de autocorrelación y correlograma [8]

La función de autocorrelación en el rezago k, denotada por ρ_k, se define como:

$$\rho_k = \frac{\gamma_k}{\gamma_o} = \frac{\text{covarianza en el rezago}}{\text{varianza}}. \qquad (2)$$

Si se grafica ρ_k respecto de k, la gráfica que se obtiene es el **correlograma poblacional.** Debido a que estos son conceptos teóricos, en la práctica tenemos la **función de autocorrelación muestral** $\hat{\rho}_k$,

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \qquad (3)$$

donde $\hat{\gamma}_k$ y $\hat{\gamma}_0$ son las covarianzas y varianza muestrales respectivamente en el rezago k, con:

$$\hat{\gamma}_k = \frac{\sum\left(Y_t - \bar{Y}\right)\left(Y_{t+k} - \bar{Y}\right)}{n}, \qquad (4)$$

$$\hat{\gamma}_0 = \frac{\sum\left(Y_t - \bar{Y}\right)^2}{n}, \qquad (5)$$

con n el tamaño de la muestra y $\bar{Y}$ la media muestral de la serie Y_t. Con lo anterior el **correlograma muestral** será entonces la gráfica de $\hat{\rho}_k$ frente a k. Una vez calculados estos valores, si el correlograma de la serie de tiempo real se parece al correlograma de una serie de tiempo de ruido blanco, se puede afirmar que quizá dicha serie es estacionaria.

6.5.5. Estadístico Q de Box – Pierce [9]

Para probar la hipótesis conjunta de que todos los ρ_k hasta ciertos rezagos son simultáneamente iguales a cero, se puede emplear el estadístico de Q de Box- Pierce definido como:

$$Q = n\sum_{i=1}^{m}\hat{\rho}_k{}^2, \qquad (6)$$

con n el tamaño de la muestra, m la longitud del rezago. El estadístico Q es común para probar si una serie de tiempo es de ruido blanco. Cuando la muestra es lo suficientemente grande, este estadístico se distribuye aproximadamente como la distribución ji cuadrada con m grados de libertad. En una aplicación, si la Q calculada excede el valor Q crítico de la distribución ji cuadrada en el nivel de significancia seleccionado, se puede rechazar la hipótesis nula de que todos los ρ_k (verdaderos) son iguales a cero; por lo menos algunos de ellos deben ser diferentes de cero.

6.5.6. Estadístico Ljung – Box [10]

Una variante del estadístico Q de Box – Pierce es el estadístico de *Ljung – Box* (LB), que se define como sigue,

$$LB = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}^2{}_k}{n-k}\right) \sim X^2{}_m, \qquad (7)$$

Aunque en muestras grandes tanto el estadístico Q como el estadístico LB siguen la distribución ji cuadrada con m grados de libertad, sin embargo se ha visto que el estadístico LB tiene mejores propiedades en muestras pequeñas (más potente, en el sentido estadístico) que el estadístico Q.

Una prueba sobre estacionariedad (o no estacionariedad) es la **prueba de la raíz unitaria**, el cual dice que un proceso estocástico lineal Y_t posee la característica de raíz unitaria si el valor de la raíz de la ecuación característica del proceso es igual a 1, en este caso se dice que el proceso es no estacionario. Si las demás raíces de la ecuación característica del proceso se encuentran dentro del círculo unitario (si r es una raíz entonces $|r| < 1$), entonces la primera diferencia del proceso es estacionaria.

6.5.7. Proceso Autoregresivo (AR) [11]

Sea Y_t un proceso estocástico definido por:

$$Y_t = c + \sum_{i=1}^{p} \varphi_i Y_{t-i} + u_t \, . \qquad (8)$$

Donde $\varphi_i, i = 1,\dots p$ son los parámetros del modelo, c es una constante y u_t es ruido blanco, decimos que Y_t sigue un modelo AR(p) (donde p es el número de términos autorregresivos) si las raíces del polinomio $Z^{p} - \sum_{i=1}^{p} \varphi_i Z^{p-1}$ están dentro del circulo unitario, es decir para cada raíz z_i se debe satisfacer $|z_i| < 1$.

6.5.8. Proceso de Medias móviles (MA)

Sea Y_t un proceso estocástico definido por:

$$Y_t = c + \beta_0 u_t + \beta_1 u_{t-1} + \ldots + \beta_q u_{t-1}, \qquad (9)$$

Donde c es una constante y u_t es ruido blanco. Aquí Y en el periodo t es igual a una constante más un promedio móvil de los términos de error (ruido blanco) presente y pasado, así en este caso decimos que sigue un proceso de medias móviles de orden q o MA(q). En otras palabras un proceso de medias móviles es una combinación lineal de términos de error de ruido blanco.

Este proceso cuenta con las siguientes propiedades: media cero, varianza mayor a la varianza del ruido, es débilmente correlacionado puesto que su autocovarianza es cero a partir de cierto valor y es además un proceso estacionario [11].

6.6.PROCESO AUTORREGRESIVO Y DE MEDIAS MOVILES (ARMA) [13]

Es un modelo que combina características de los modelos AR y MA, para un proceso estocástico Y_t definido como sigue:

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i u_{t-1} + u_t, \qquad (10)$$

Donde c es una constante y u_t es ruido blanco. Se dice que el proceso Y_t satisface un proceso ARMA (p,q).

6.7.PROCESO AUTORREGRESIVO INTEGRADO DE MEDIAS MÓVILES (ARIMA) [14]

Algunos modelos de series de tiempo se basan en el supuesto de que las series de tiempo consideradas son estacionarias, pero se sabe que muchas series de tiempo económicas no son estacionarias, es decir son integradas. Luego para hacer una serie de tiempo estacionaria se debe diferenciar d veces y luego aplicarle el modelo ARMA (p,q), por ello se dice que la serie de tiempo original es ARIMA (p,d,q), es decir una serie de tiempo autorregresiva integrada de medias móviles, donde p nota el número de términos autorregresivos, d el número de veces que se debe diferenciar la serie para hacerse estacionaria y q el número de términos de medias móviles. Note que si $d=0$ el modelo ARIMA (p,d,q) se convierte en un modelo ARMA (p,q).

7. ANÁLISIS Y PLANTEAMIENTO DE LA METODOLOGÍA

En la Figura 3. se puede observar los boxplot (diagrama de cajas y bigotes) de las series de tiempo a tratar. Con esto es posible evidenciar que los datos no son simétricos (existe algún grado de sesgo), no se presentan valores atípicos (outliers) que puedan afecten notablemente los datos. En los tres casos (diagramas de caja y bigote) se tiene que los datos tienen una tendencia (un sesgo leve) hacia el primer cuartil, (es decir que los valores son bajos)manteniendo bajos precios

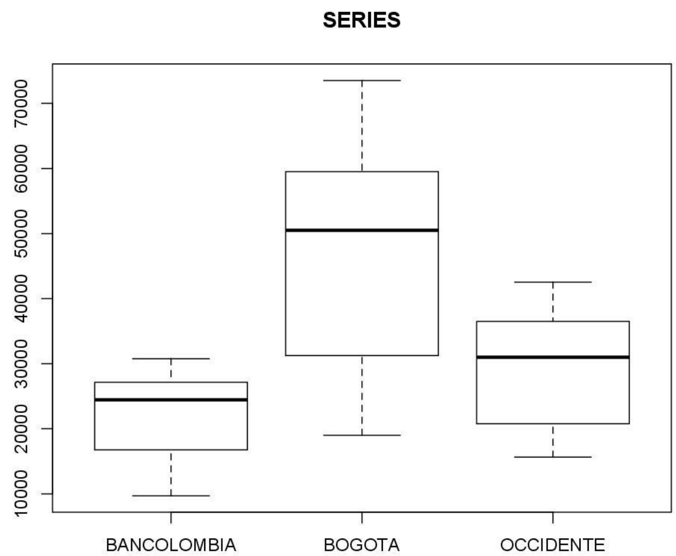

Figura 3. BOXPLOT de los históricos del precio de cierre de los activos de los bancos Bancolombia, Bogotá y Occidente.

Se realiza un análisis de la serie del activo Bancolombia para un periodo de tiempo comprendido en los primeros dos meses del año 2006, esto con el fin de presentar un método de predicción de activos haciendo uso del programa R.

En la Figura 4. se presenta el histórico del precio de cierre del activo Bancolombia, que tiene una tendencia creciente no estricta, se puede observar que la media no es estacionaria y la varianza cambia en los instantes de tiempo donde se presentan los picos.

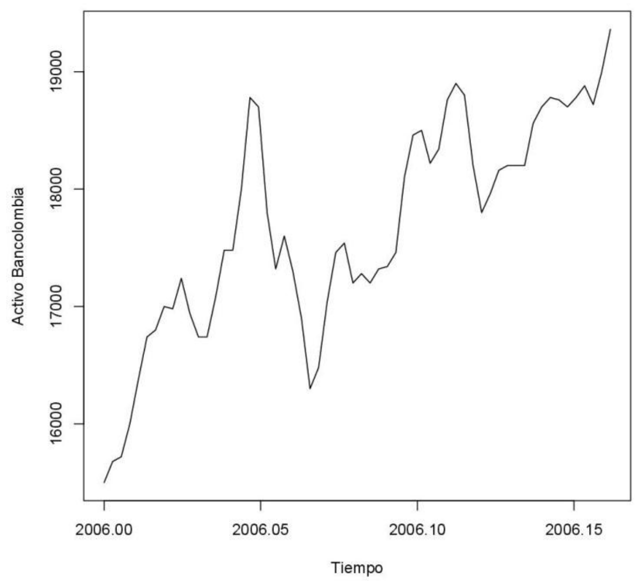

Figura 4. Serie de tiempo del activo Bancolombia en los dos primeros meses del año 2006

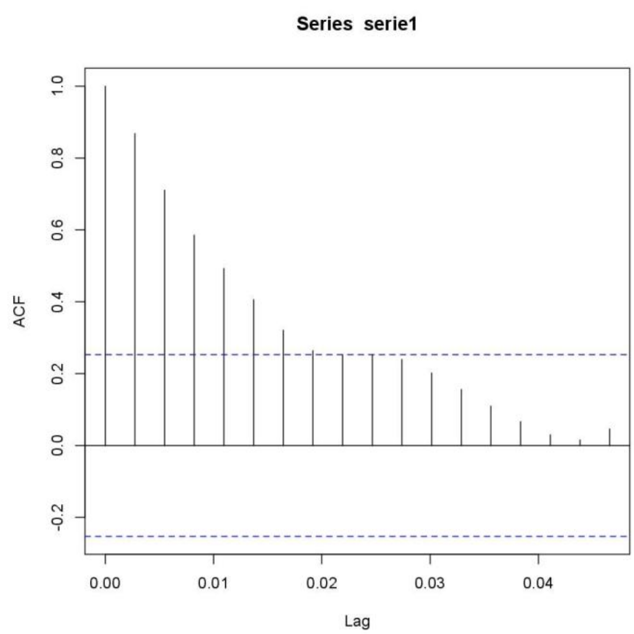

Figura 5. Función de Autocorrelación (ACF) del activo Banco Bancolombia

La Figura 5. presenta la función de autocorrelación surge de diferenciar la serie una vez. Se ve que tiene un decaimiento lento y tiende hacia a cero, y la función de autocorrelación parcial presenta un valor cercano a uno en el primer rezago y el resto es aproximadamente cero. En R la prueba de Dickey-Fuller arroja que la serie es de raíz unitaria, entonces tal como se ve en [1], el proceso tiene la forma de un ARIMA(p,1,q), en este caso ARIMA(2,1,2).

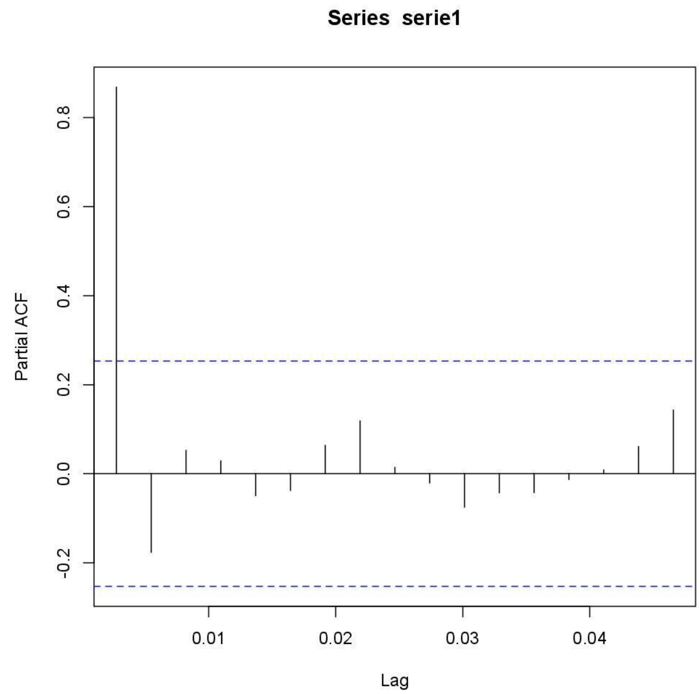

Figura 6. Función de Autocorrelación parcial (Partial ACF) del activo Banco Bancolombia

La figura 6. muestra la función de autocorrelación parcial del activo (presentado una muy rápida convergencia, Después del primer rezago converge hacia cero sin salirse de sus bandas, indicando un MA(2).

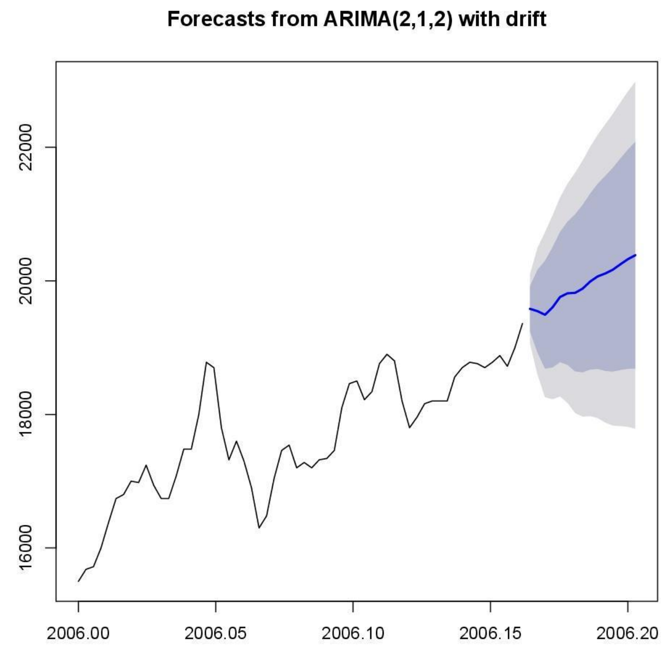

Figura 7. Predicción del activo Banco Bancolombia con un modelo ARIMA(2,1,2) en 15 días posteriores a los datos simulados.

En la figura 7. se puede observar la predicción del activo con un modelo ARIMA(2,1,2), obtenido mediante la función auto.arima del programa R, mostrando que que la serie de los activos durante este periodo de tiempo fue bien comportada y presentando una posible "buena predicción" de los datos obtenidos, la cual se estudiará a partir de la tabla 1.

Valor Real	Predicción	Límite Inferior del Intervalo de Confianza 95%	Límite Superior del Intervalo de Confianza 95%	Error Absoluto |Valor Real - Predicción|
19400	19580.37	19064.27	20096.48	180.37
19560	19545.13	18596.65	20493.60	14.87
20120	19491.47	18256.65	20726.28	628.53

20300	19602.54	18227.23	20977.85	697.46
20000	19757.62	18267.69	21247.56	242.38
19800	19812.77	18171.39	21454.16	12.77
18900	19819.60	18022.22	21616.99	919.6
18860	19882.87	17966.48	21799.26	1022.87
19200	19988.88	17963.79	22003.96	788.88
19420	20066.00	17945.80	22186.20	646.00
19360	20109.39	17878.52	22340.27	749.39
19500	20165.27	17834.03	22496.52	665.27
19480	20245.77	17825.47	22666.08	765.77
19460	20322.82	17814.97	22830.68	862.82
19720	20382.99	17785.82	22980.17	662.99

Tabla 1. Análisis de error de predicción del activo banco Bancolombia.

A partir de los datos de la tabla 1. se puede observar que los menores errores absolutos en la predicción se presentan en la predicción hasta el día 6 posterior a los datos que se obtienen, siendo este error no superior al 3.44% del valor real, el cual se encuentra dentro del intervalo de confianza obtenido. Posterior a este dato se encuentra un incremento en los errores siendo de hasta un 5.43% del valor real, presentando un alto grado de error no admisible, pese a que el valor real del activo cae sobre el intervalo de confianza. Con esto se puede decir que la metodología propuesta para la predicción de este activo puede llegar a ser eficiente hasta la primera semana de la predicción.

El mismo procedimiento es empleado para el Banco de Bogotá y el Banco de Occidente, omitiendo el análisis de los resultados y solo presentando los resultados numéricos obtenidos.

A continuación se presentan los resultados obtenidos para el activo Banco de Bogotá.

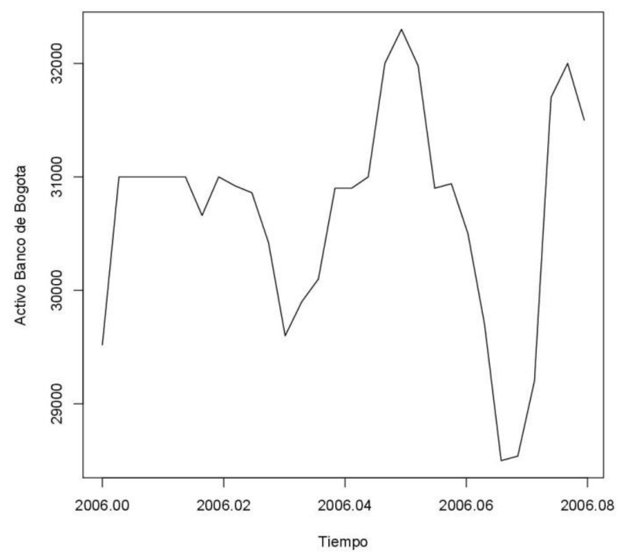

Figura 8. Serie de tiempo del activo Banco de Bogotá en el primer mes del año 2006.

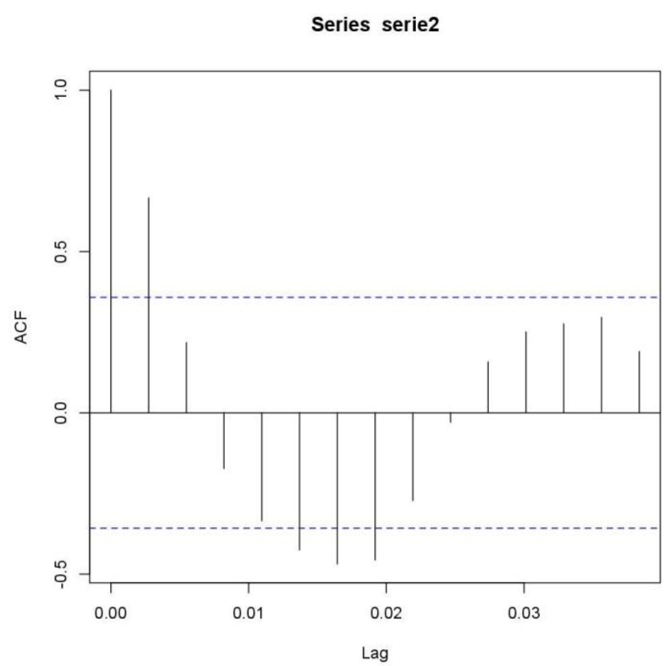

Figura 9. Función de Autocorrelación (ACF) del activo Banco de Bogotá.

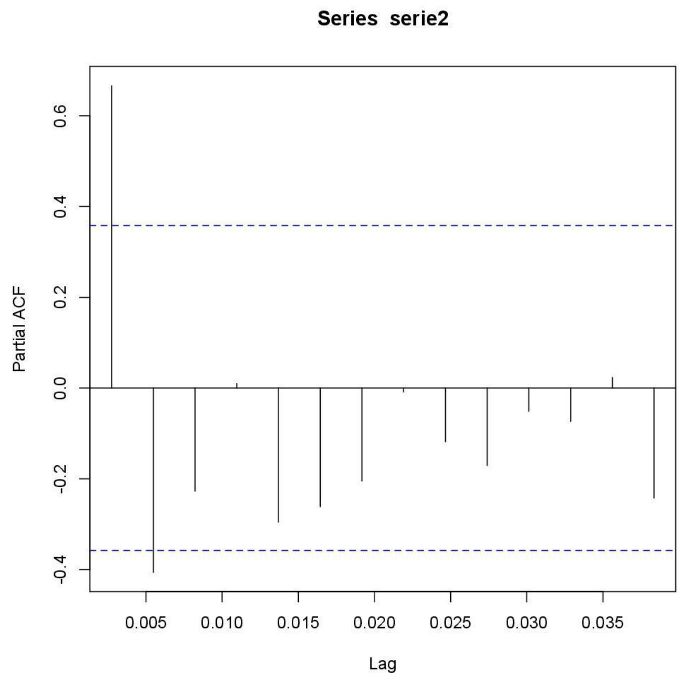

Figura 10. Función de Autocorrelación parcial (Partial ACF) del activo Banco de Bogotá.

Al hacer uso del autoarima R arroja que el modelo que mejor se ajusta a la serie del Banco Bogotá es un ARIMA(2,0,0), es decir no tiene que ser diferenciado para ser estacionario, algo que se ve al mirar las raíces del proceso, y que dan fuera del círculo unitario.

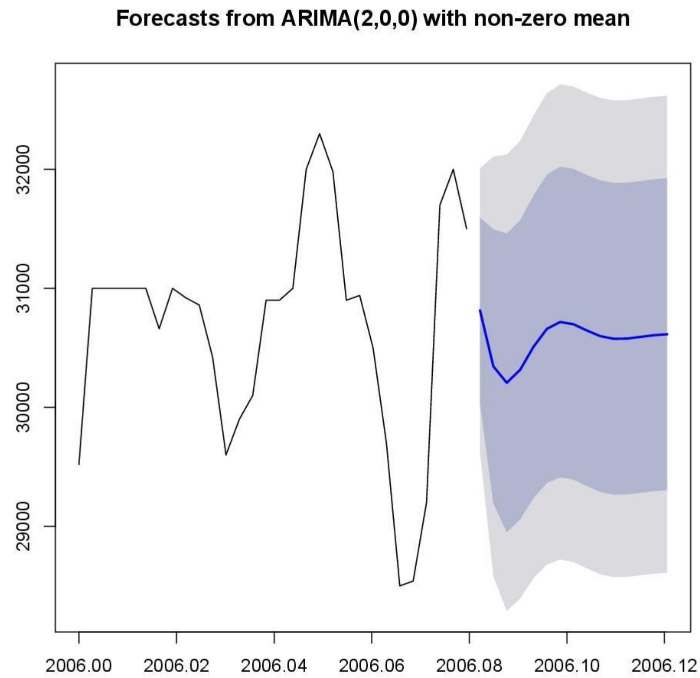

Figura 11. Predicción del activo Banco de Bogotá con un modelo ARIMA(2,0,0) en 15 días posteriores a los datos simulados.

Valor Real	Predicción	Límite Inferior del Intervalo de Confianza 95%	Límite Superior del Intervalo de Confianza 95%	Error Absoluto |Valor Real - Predicción|
31200	30815.13	29622.65	32007.60	384.87
30800	30343.98	28582.02	23105.93	456.02
30800	30206.55	28288.51	32124.58	593.45
31000	30315.08	28393.66	32236.50	684.92
31040	30508.39	28565.06	32451.72	531.61
32000	30659.23	28679.29	32639.17	1340.77
32520	30717.44	28721.42	32713.46	1802.56
32700	30698.15	28700.99	32695.31	2001.85

32780	30645.29	28647.07	32643.52	2131.71
32520	30598.36	28597.24	32599.48	1921.64
32840	30576.26	28537.47	32597.04	2263.74
33500	30577.92	28574.91	32580.92	2922.08
33840	30591.82	28588.78	32594.87	3248.18
32000	30606.05	28602.78	32609.31	1393.95
32000	30613.90	28610.47	32617.32	1386.10

Tabla 2.Análisis de error de predicción del activo banco de Bogotá..

A continuación se realiza el análisis para el activo Banco de Occidente.

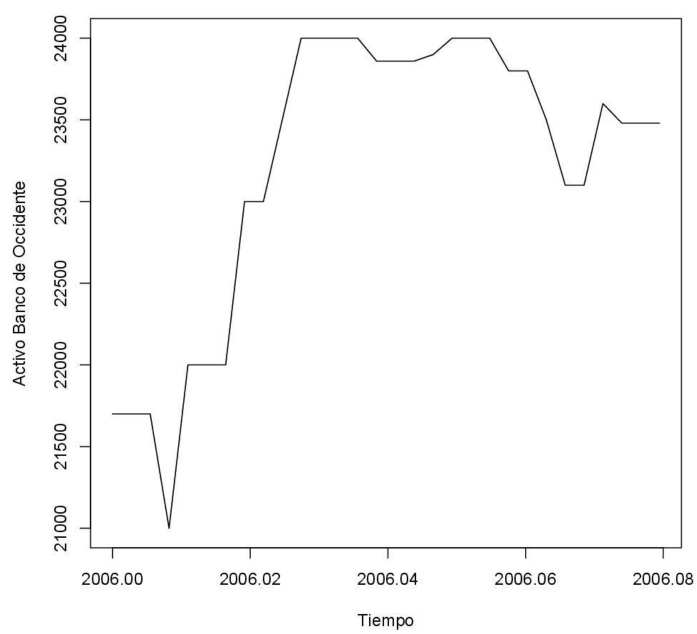

Figura 12. Serie de tiempo del activo Banco de Occidente en el primer mes del año 2006

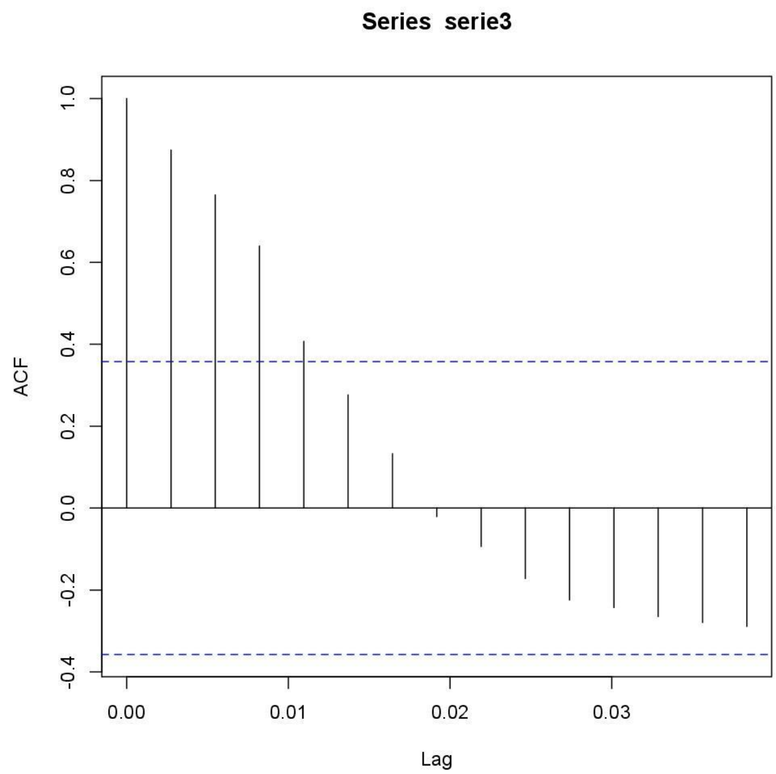

Figura 13. Función de Autocorrelación (ACF) del activo Banco de Occidente.

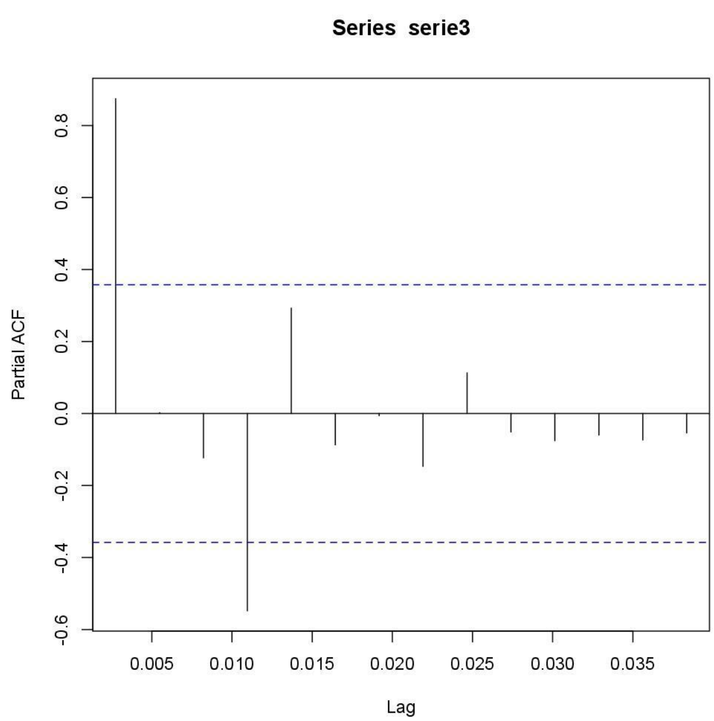

Figura 14. Función de Autocorrelación parcial (Partial ACF) del activo Banco de Occidente.

Al hacer uso del autoarima R arroja que el modelo que mejor se ajusta a la serie del Banco Occidente es un ARIMA(0,1,0), es decir un modelo no autorregresivo de media móvil 0 y una única diferenciación.

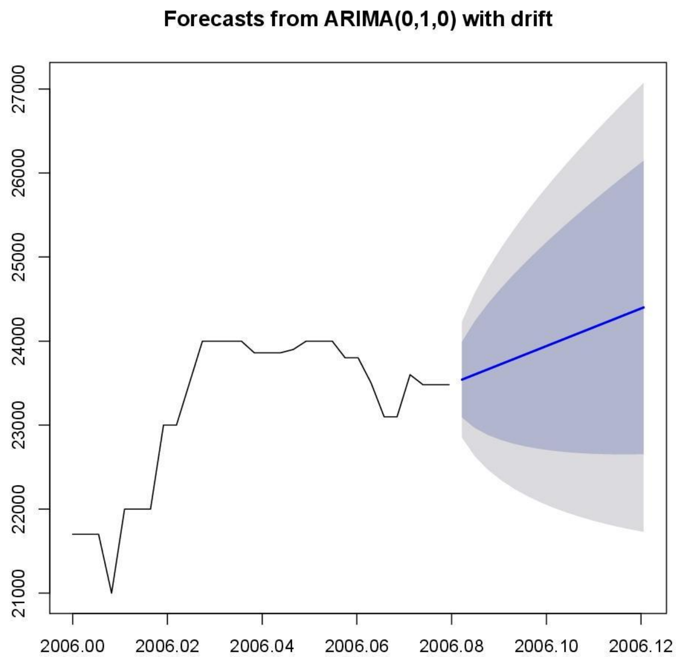

Figura 15. Predicción del activo Banco de Occidente con un modelo ARIMA(0,1,0) en 15 días posteriores a los datos simulados.

Valor Real	Predicción	Límite Inferior del Intervalo de Confianza 95%	Límite Superior del Intervalo de Confianza 95%	Error Absoluto |Valor Real - Predicción|
23480	23541.38	22851.72	24231.04	61.38
23480	23602.76	22627.43	24578.08	122.76
23480	23664.14	22469.61	24858.66	184.14
23000	23725.52	22346.20	25104.83	725.52
23000	23786.90	22244.77	25329.02	786.90

23000	23848.28	22158.96	25537.59	848.28
23000	23909.66	22084.99	25734.32	909.66
24000	23971.03	22020.38	25921.68	28.97
24000	24032.41	21963.44	26101.39	32.41
23500	24093.79	21912.90	26274.69	593.79
24000	24155.17	21867.83	26442.51	155.17
24000	24216.55	21827.50	26605.60	216.55
24000	24277.93	21791.33	26764.53	277.93
23800	24339.31	21758.84	26919.78	539.31
23800	24400.69	21729.65	27071.73	600.69

Tabla 3. Análisis de error de predicción del activo banco de Occidente..

Con base en la metodología propuesta en el documento se realiza la simulación para la predicción de los activos del precio de cierre de los bancos Bancolombia, Bogotá y Occidente. La predicción se realiza con proyección a la primera semana del mes de agosto, tomando con base de la simulación el mes de julio del presente año (2017). A continuación se realiza el análisis de los resultados obtenidos en dicha simulación para cada uno de los tres bancos mencionados.

8. ANÁLISIS Y PREDICCIÓN DE LOS VALORES FUTUROS

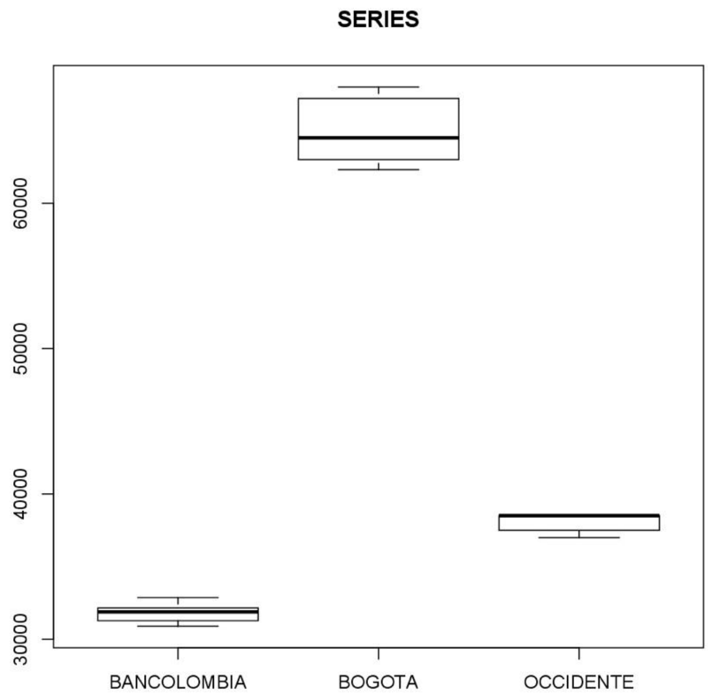

Figura 16. Boxplot de las series

En la Figura 16 del diagrama de cajas podemos ver los valores máximos y mínimos que toman las series de los bancos Bancolombia, Bogotá y Occidente. Se ve además que no presentan valores atípicos y que la distribución de los datos en los dos primeros bancos es de forma casi simétrica mientras que en el banco de occidente la información se encuentra en los primeros tres cuartiles con la mediana el valor máximo de la serie.

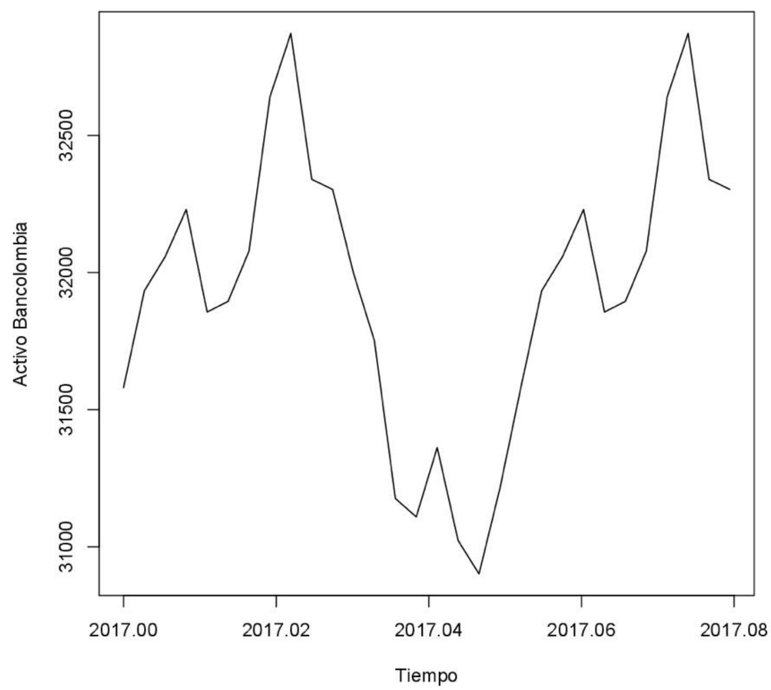

Figura 17. Serie Activo Bancolombia

En la Figura 17 podemos observar como es el comportamiento de activo Bancolombia en el mes de julio, representando dos picos máximos y un decaimiento, correspondiente al 6,2% del valor promedio que toma el activo durante este mes. No es una serie con comportamiento creciente o decreciente, lo cual dificulta la predicción de esta al igual que no es estable en cuanto a que ésta varíe con puntos cercanos de acuerdo a un valor medio.

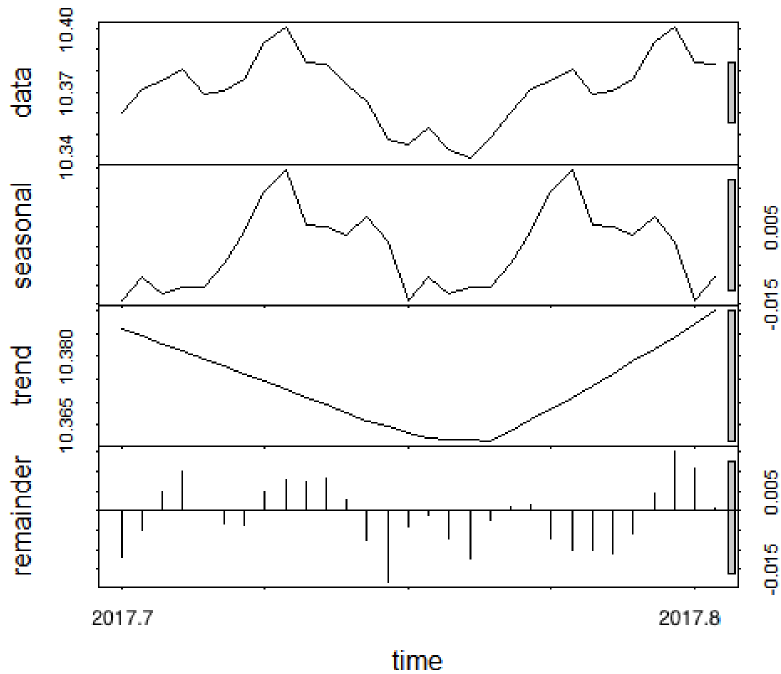

Figura 17.1. Descomposición estacional usando la función stl() del activo Bancolombia.

En la figura 17.1 podemos observar los datos de la serie (data) que es la serie de tiempo del activo Bancolombia, podemos ver que esta al no ser una serie "bien comportada" la estacionalidad de la misma es muy parecida a la serie de los datos, en la tendencia (trend) podemos observar que esta serie no es creciente ni decreciente, por ello es más complicada de simular o predecir por métodos de regresión simple o compuesta. En el residuo (remainder) se puede verificar que estos son positivos y negativos en diferentes intervalos por la misma tendencia que presenta esta serie.

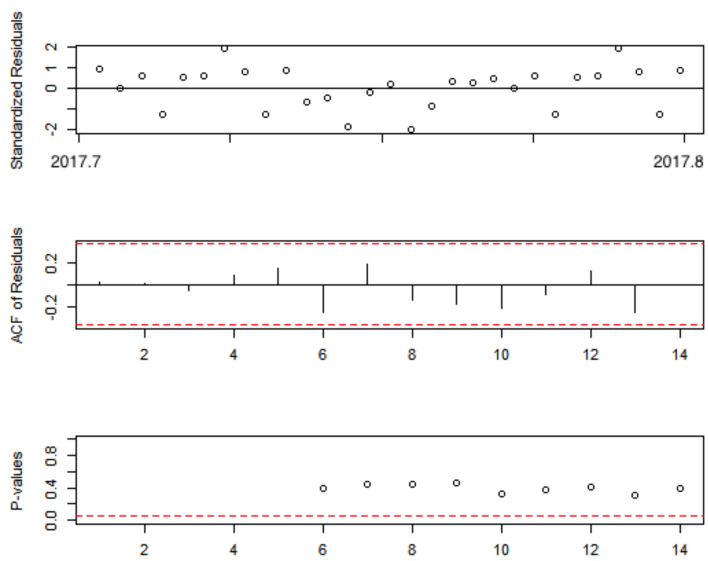

Figura 17.2. Grafico de Residuales del Banco Bancolombia

El primer paso para verificar el diagnostico de nuestro modelo ajustado es analizar los residuales del ajuste para asegurar que no haya dependencia. Las gráficas anteriores con las que se determina este análisis se obtiene a partir del comando tsdiag(), donde se obtienen los p-values de la estadística de Ljung-Box para los primeros 14 rezagos. Se puede ver que en la ACF de los residuales los valores son cercanos a cero, indicando que los residuales muy posiblemente son ruido blanco y por lo tanto independientes e idénticamente distribuidos unos de otros. Además, los p-values después del sexto rezago superan el nivel de significancia 0.05 que viene por defecto, indicando que no hay suficiente evidencia para rechazar que son independientes, que es lo se necesita para un buen ajuste del modelo.

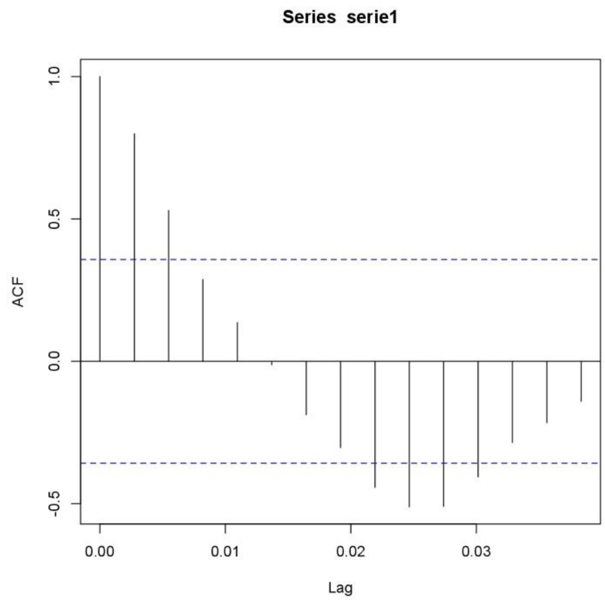

Figura 18. ACF serie Bancolombia

La función de autocorrelación para la serie de Bancolombia dado que tiene picos grandes alternados con valores negativos y positivos que salen de las bandas de confianza indica que hay correlación, lo que sugiere una parte autorregresiva del proceso sin parte de medias móviles, que con la función de autocorrelación parcial se puede determinar su orden.

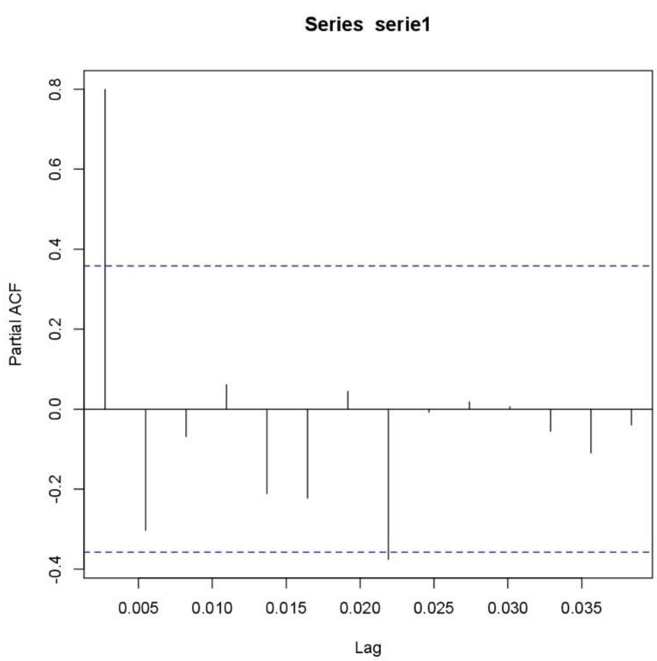

Figura 19. PACF serie Bancolombia

Para el activo Bancolombia mediante el comando Box.test() obtenemos la estadística de prueba de esta serie, teniendo como p valor 0.9831 lo que muestra y corrobora con lo hecho anteriormente, que no hay evidencia suficiente para ver que los residuales son dependientes, resultado favorable para el ajuste de esta serie en nuestro trabajo.

Al analizar la función de autocorrelación parcial de la serie Bancolombia tiene dos valores significativos que salen de la banda de confianza, más específicamente el primero positivo de mayor longitud respecto al segundo de valor negativo. Esto indica que el proceso es autorregresivo de orden 2, tal como lo predice la función en R.

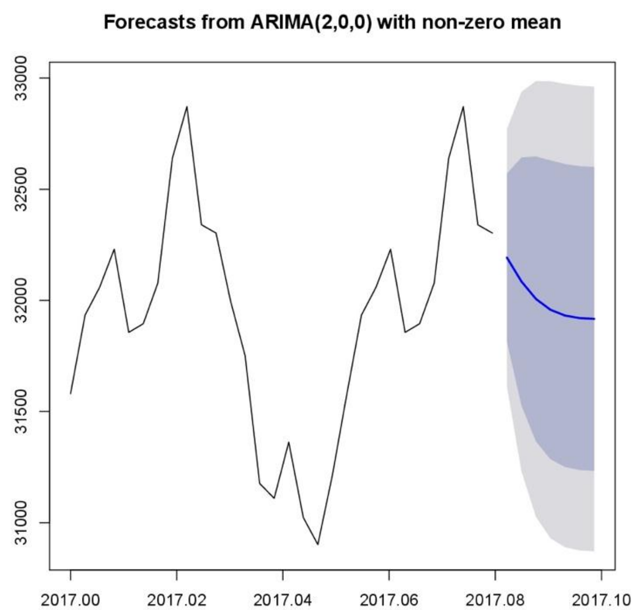

Figura 20. Predicción serie Bancolombia

En la Figura 20 se presenta la predicción del activo Bancolombia mediante un modelo ARIMA(2,0,0), lo que lo hace una serie autorregresiva de media cero sin diferenciaciones, en el gráfico podemos ver que la predicción es decreciente durante la semana, cuyos valores se presentan en la Tabla 4.

Día	Predicción	Límite inferior del Intervalo de confianza al 95%	Límite superior del Intervalo de confianza al 95%
1	32192.26	31613.73	32770.79
2	32084.78	31231.21	32938.34
3	32006.05	31025.32	32986.78
4	31957.41	30929.06	32985.76
5	31931.57	30889.20	32973.93
6	31920.16	30874.81	32965.52
7	31916.63	30870.92	32962.35

Tabla 4. Predicción de los valores del activo Bancolombia para la primera semana de Agosto del año 2017.

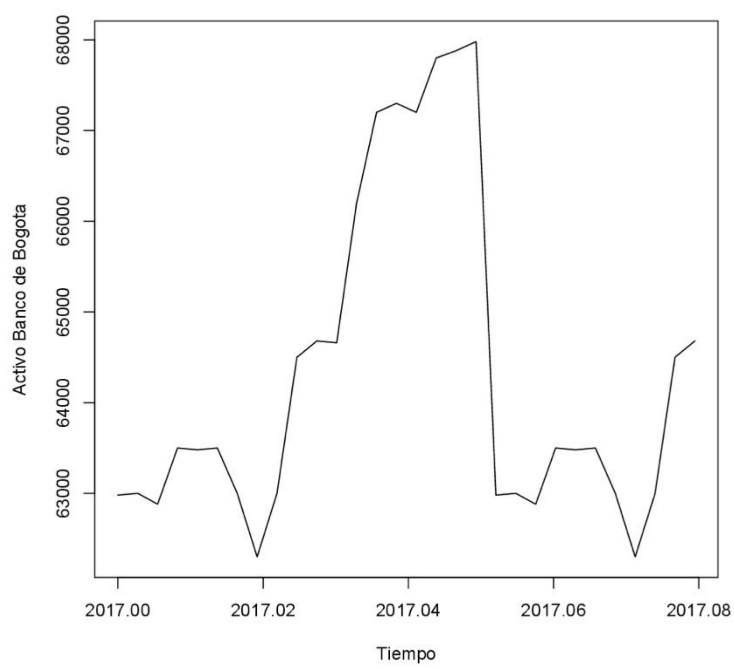

Figura 21. Serie activo banco de Bogotá

En la Figura 21 se muestra la serie de tiempo del activo Bogotá, mostrando que tiende a ser creciente en algunos periodos de tiempo e intentando mantener la misma tendencia durante esos periodos. Podemos observar que el activo Bogotá a diferencia del activo Bancolombia este presenta una variación mayor en los picos máximos y mínimos con respecto al valor medio, esta variación llega a ser de hasta el 8,8% en el mes.

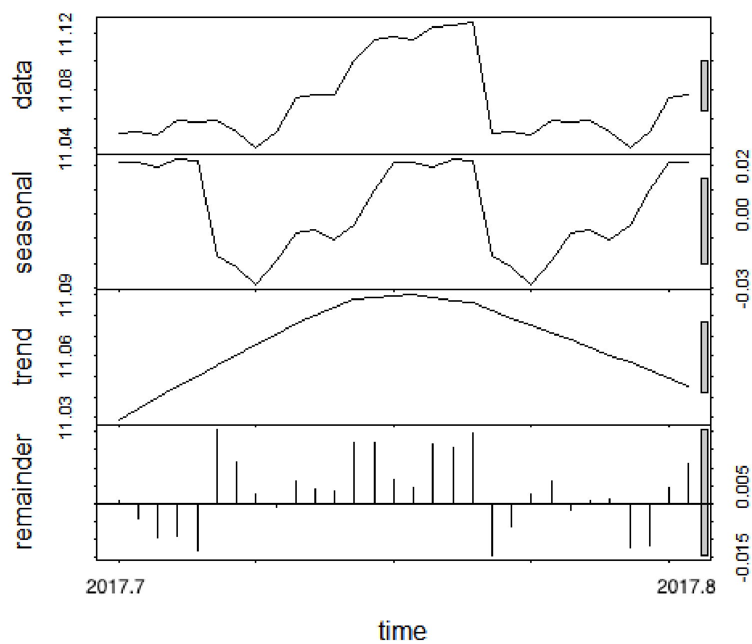

Figura 21.1. Descomposición estacional usando la función stl() del activo Banco de Bogotá

En la figura 21.1 podemos observar los datos de la serie (data) que es la serie de tiempo del activo Banco de Bogotá, podemos ver que esta al no ser una serie “bien comportada” la estacionalidad de la misma es muy parecida a la serie de los datos, en la tendencia (trend) podemos observar que esta serie no es creciente ni decreciente, pero en este intervalo de tiempo tiene una tendencia parabólica negativa, con un máximo absoluto en casi la mitad del tiempo. En el residuo (remainder) se puede verificar que estos son positivos en su gran mayoría por la tendencia parabólica de la misma.

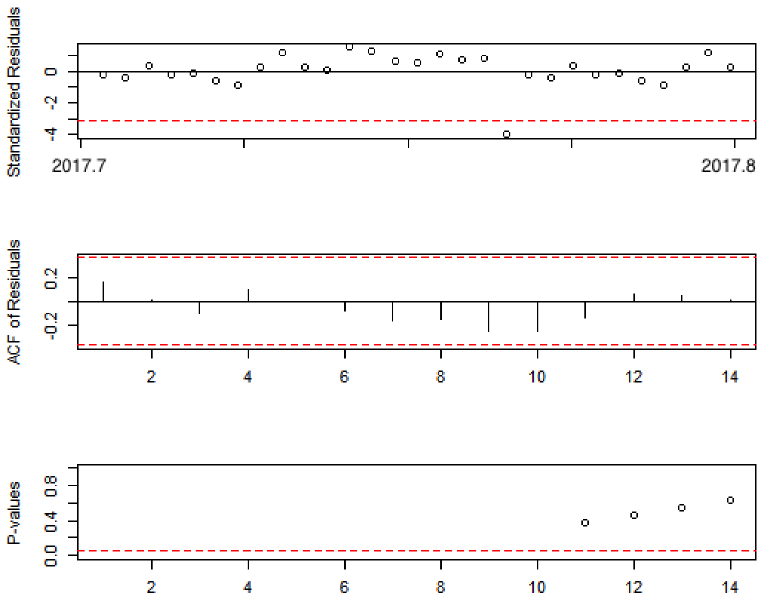

Figura 21.2. Grafico de residuales Banco de Bogotá

Se puede ver que en la ACF de los residuales los valores son cercanos a cero, indicando muy posiblemente son ruido blanco y por lo tanto independientes e idénticamente distribuidos unos de otros. Además, los p-values después del décimo primer rezago superan el nivel de significancia 0.05 que viene por defecto, indicando que no hay suficiente evidencia para rechazar que son independientes, que es lo se necesita para un buen ajuste del modelo.

Análogamente como en el activo de Bancolombia, obtenemos la estadística de prueba de esta serie, teniendo como p valor 0.3569, lo que muestra y corrobora con lo hecho anteriormente, que no hay evidencia suficiente para ver que los residuales son dependientes.

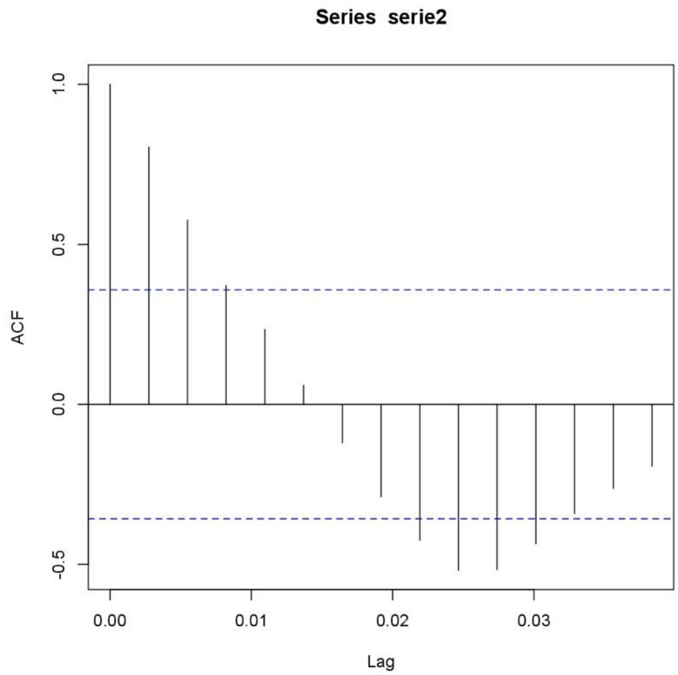

Figura 22. ACF serie banco de Bogotá

Similar a como sucede con la serie del banco Bancolombia, el comportamiento de la función de autocorrelación tiene valores significativos tanto positivos como negativos en forma de onda que salen de la banda de confianza, lo que sugiere que puede ser modelado mediante un proceso autorregresivo puro. El orden de la parte autorregresiva de nuevo se puede determinar analizando la función de autocorrelación parcial.

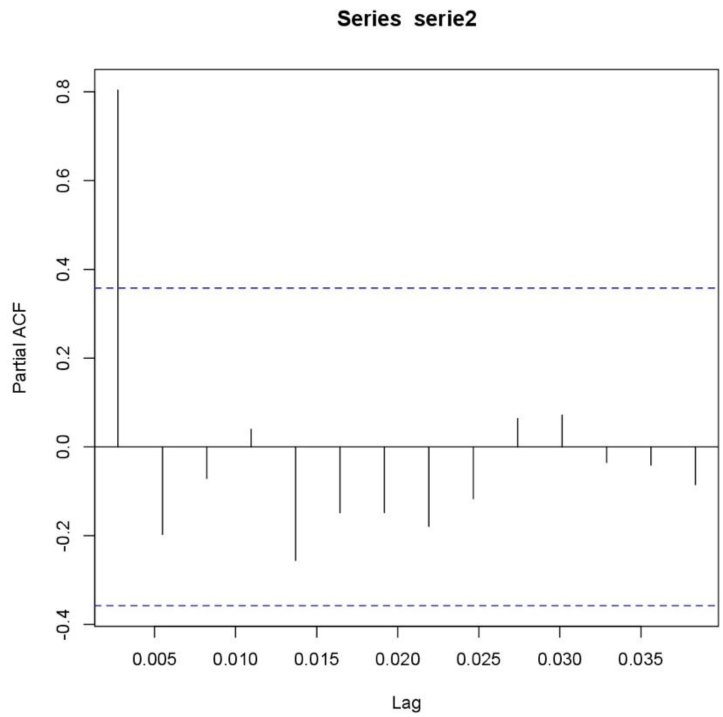

Figura 23. PACF serie banco de Bogotá

Dado que la función de autocorrelación parcial para esta serie solamente tiene un valor positivo que sobresale fuera de la banda de confianza y el resto de valores se estabilizan dentro de ella, esto sugiere que el orden del modelo autorregresivo sea de orden 1.

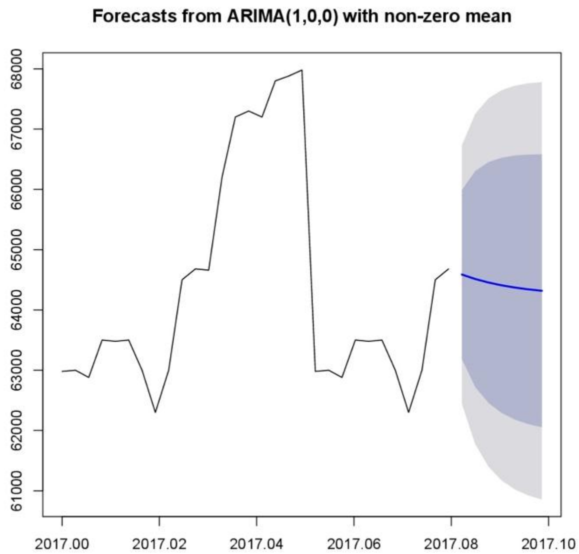

Figura 24. Predicción serie banco de Bogotá

En la Figura 24 se presenta la predicción del activo Bogotá mediante un modelo ARIMA(1,0,0). Podemos ver que la predicción durante la semana es decreciente casi constante, en la Tabla 5, se presenta la simulación del valor futuro, donde se puede ver que se espera un decaimiento, no obstante es una predicción que depende de los límites de confianza y podrían llegar a haber algunos días de la semana en los que se presente un incremento y no un decaimiento del activo.

Día	Predicción	Límite inferior del Intervalo de confianza al 95%	Límite superior del Intervalo de confianza al 95%
1	64586.93	62442.65	66731.22
2	64512.96	61773.88	67252.05
3	64454.17	61398.44	67509.90
4	64407.44	61167.59	67647.30
5	64370.30	61019.35	67721.26

6	64340.78	60921.50	67760.07
7	64317.32	60855.57	67779.07

Tabla 5. Predicción de los valores del activo Bogotá para la primera semana de Agosto del año 2017.

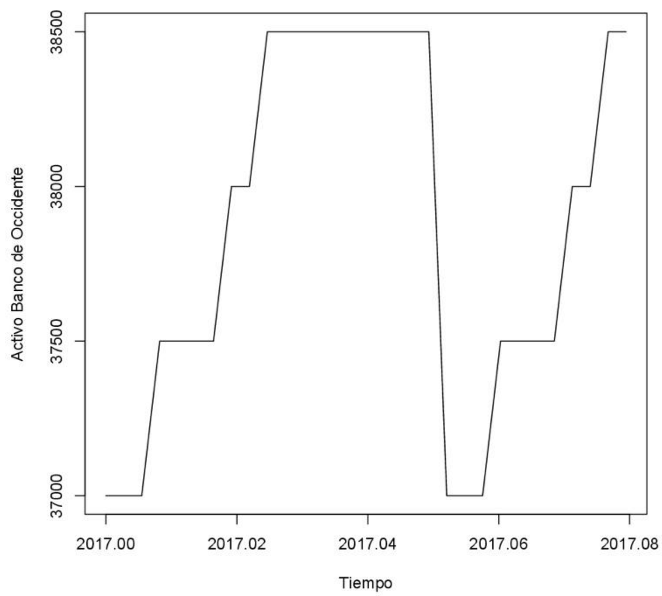

Figura 25. Serie activo banco de Occidente

En la Figura 25 se muestra la serie de tiempo del activo Occidente con un crecimiento en un primer periodo de tiempo, presentando posteriormente un decaimiento no tan significativo (3.9% con respecto al valor medio).

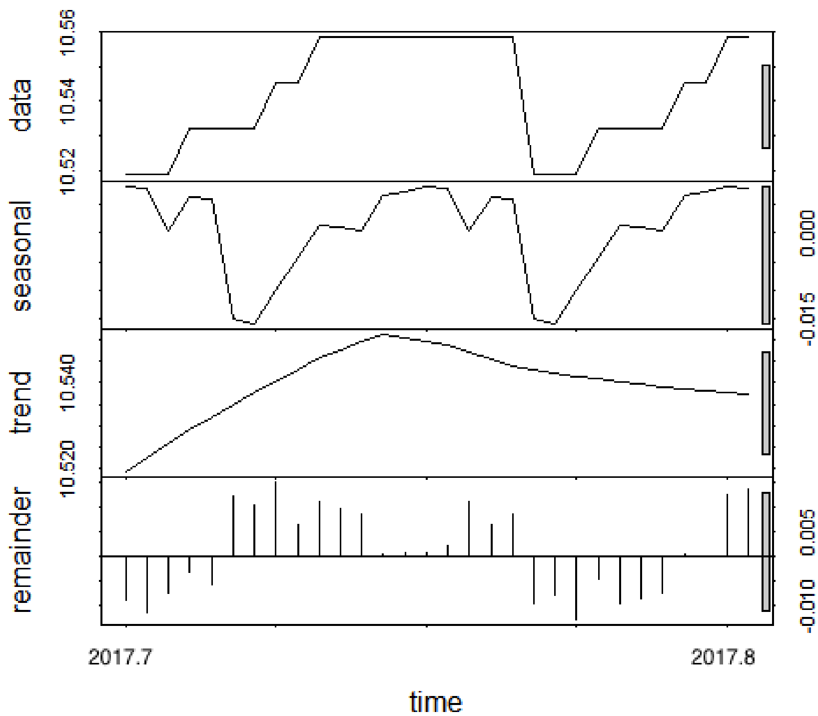

Figura 25.1 Descomposición estacional usando la función stl() del activo banco de occidente

En la figura 25.1 podemos observar los datos de la serie (data) que es la serie de tiempo del activo Bancolombia, podemos ver que esta al no ser una serie "bien comportada" la estacionalidad de la misma es muy parecida a la serie de los datos, en la tendencia (trend) podemos observar que esta serie no es creciente ni decreciente, pero a diferencia de los otros dos activos esta serie no decrece de una forma tan rápida y por el contrario se mantiene constante en ciertos puntos, por ello la serie data parece una escalera. En el residuo (remainder) se puede verificar que estos son positivos y negativos en diferentes intervalos por la misma tendencia que presenta esta serie, al igual que los residuos no son todos crecientes sino que intentan mantenerse casi constantes en los intervalos en que cambian.

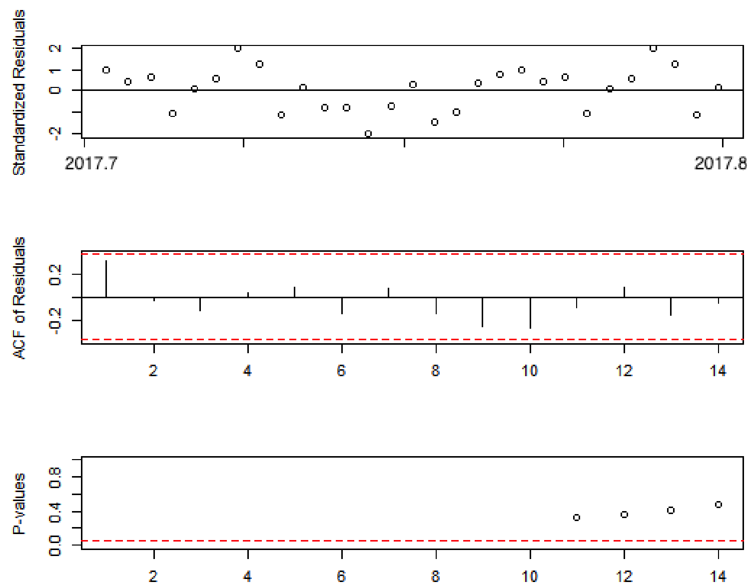

Figura 25.2 Grafico de Residuales de Banco Occidente

Se puede ver que en la ACF de los residuales los valores son cercanos a cero, indicando muy posiblemente son ruido blanco y por lo tanto independientes e idénticamente distribuidos unos de otros. Su comportamiento es muy parecido al anterior activo (Bogotá), es decir, los p-values después del décimo primer rezago superan el nivel de significancia 0.05 que viene por defecto, indicando que no hay suficiente evidencia para rechazar que son independientes, que es lo se necesita para un buen ajuste del modelo.

Así como en los casos anteriores, obtenemos la estadística de prueba de esta serie, teniendo como p valor 0.7796 lo que muestra que no hay evidencia suficiente para ver que los residuales son dependientes. Podemos concluir que en las tres series obtuvimos que los residuales muy posiblemente son independientes, algo fundamental en la verificación de supuestos en la teoría de series.

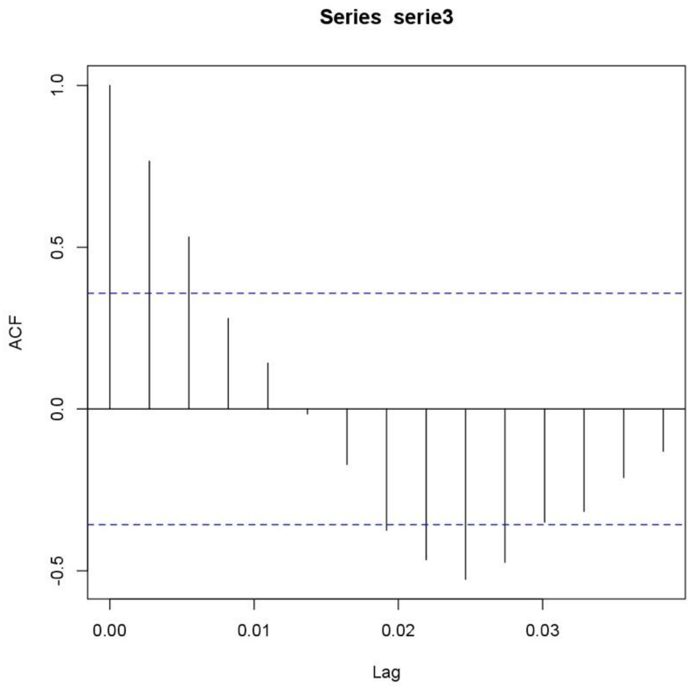

Figura 26. ACF serie activo banco de Occidente

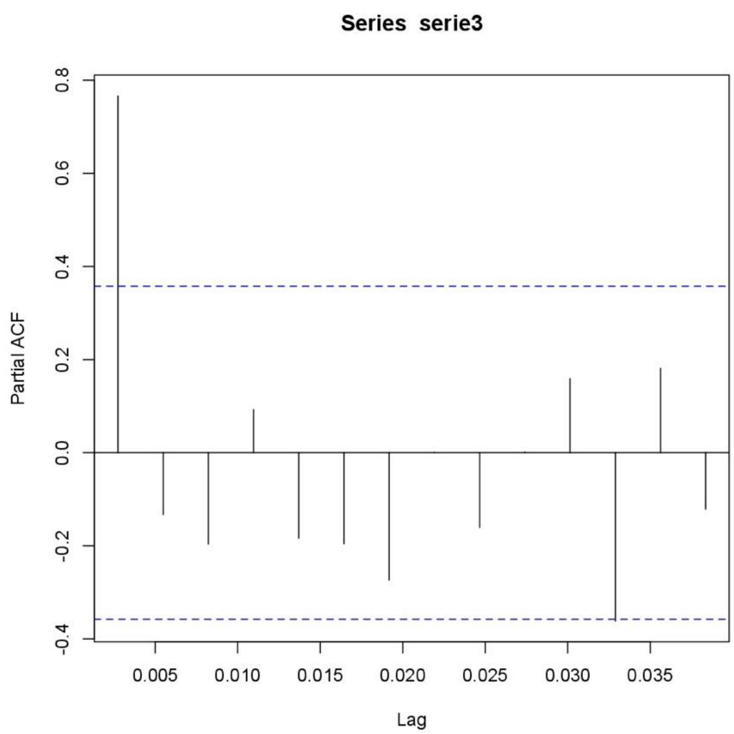

Figura 27. PACF serie activo banco de Occidente

En la serie obtenida por el banco de Occidente, al analizar el comportamiento de las funciones de autocorrelación simple y parcial, exactamente pasa igual que con la serie del banco de Bogotá, es decir, que puede ser vista como un proceso autorregresivo de orden 1, lo que coincide con los resultados arrojados con el programa en R.

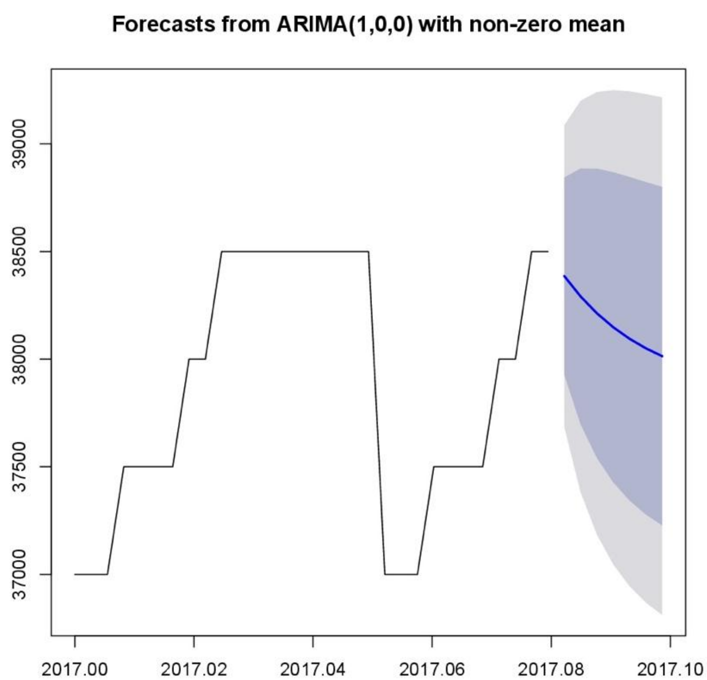

Figura 28. Predicción serie activo banco de Occidente

En la Figura 28 se presenta la predicción del activo Occidente mediante un modelo ARIMA(1,0,0), es un modelo autorregresivo sin diferenciación y de media 0. Al igual que para los activos anteriores se presenta una predicción decreciente durante la primera semana del mes de agosto. En la Tabla 6 se presenta la predicción generada para la primera semana del mes de agosto de 2017. Tomando valores decrecientes en cada periodo de tiempo, sin embargo se espera que el valor real que pueda tomar el activo no supere los umbrales de los límites de confianza.

Día	Predicción	Límite inferior del Intervalo de confianza al 95%	Límite superior del Intervalo de confianza al 95%
1	38385.61	37684.94	39086.27
2	38290.99	37381.73	39200.26
3	38212.74	37184.89	39240.59
4	38148.02	37046.38	39249.67
5	38094.50	36945.10	39243.89
6	38050.22	36869.28	39231.17
7	38013.61	36811.55	39215.66

Tabla 6. Predicción de los valores del activo Occidente para la primera semana de Agosto del año 2017.

9. PRUEBAS DE NORMALIDAD PARA LOS RESIDUALES

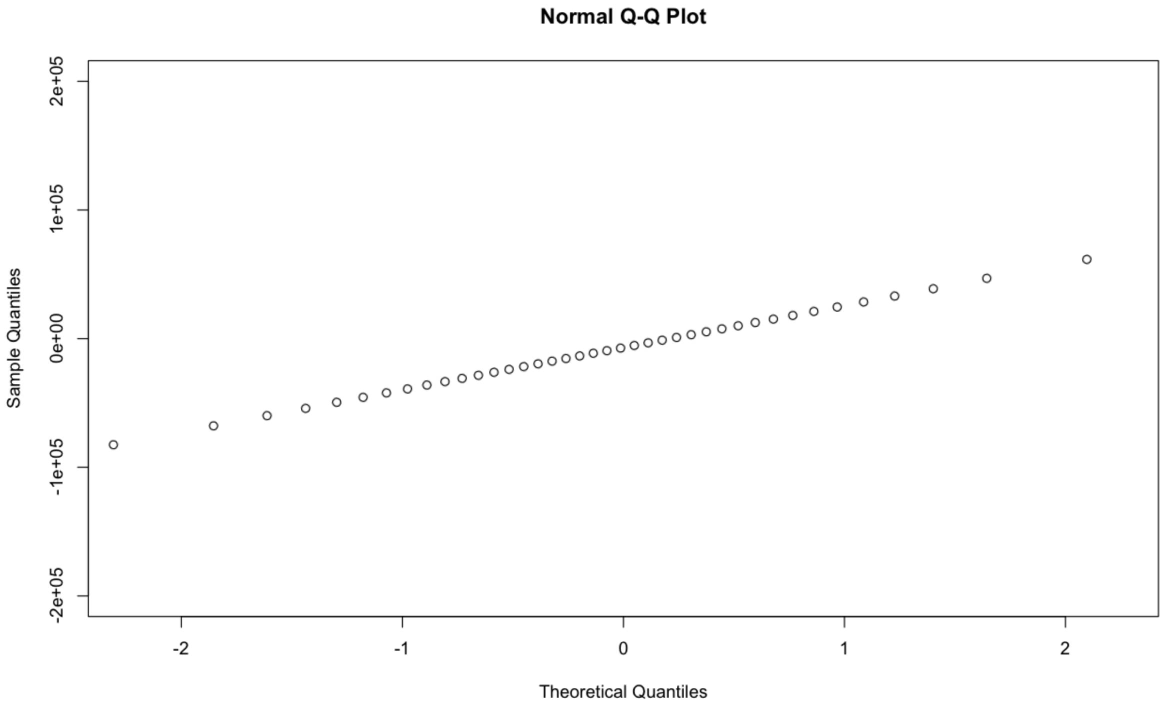

Figura 29. QQ plot de la predicción Bancolombia.

En la figura 29 se evaluará la normalidad de los residuales de la serie Bancolombia. Se puede observar en el anterior gráfico que al comparar los cuantiles de la distribución normal estándar con los de los residuales se tiene un comportamiento lineal, lo cual es un buen indicio. Se realiza el test de Jarque- Bera y se obtiene la estadística de valor 0.31546 con 2 grados de libertad y un p-value = 0.8541 donde confirmamos que los residuales se distribuyen normalmente.

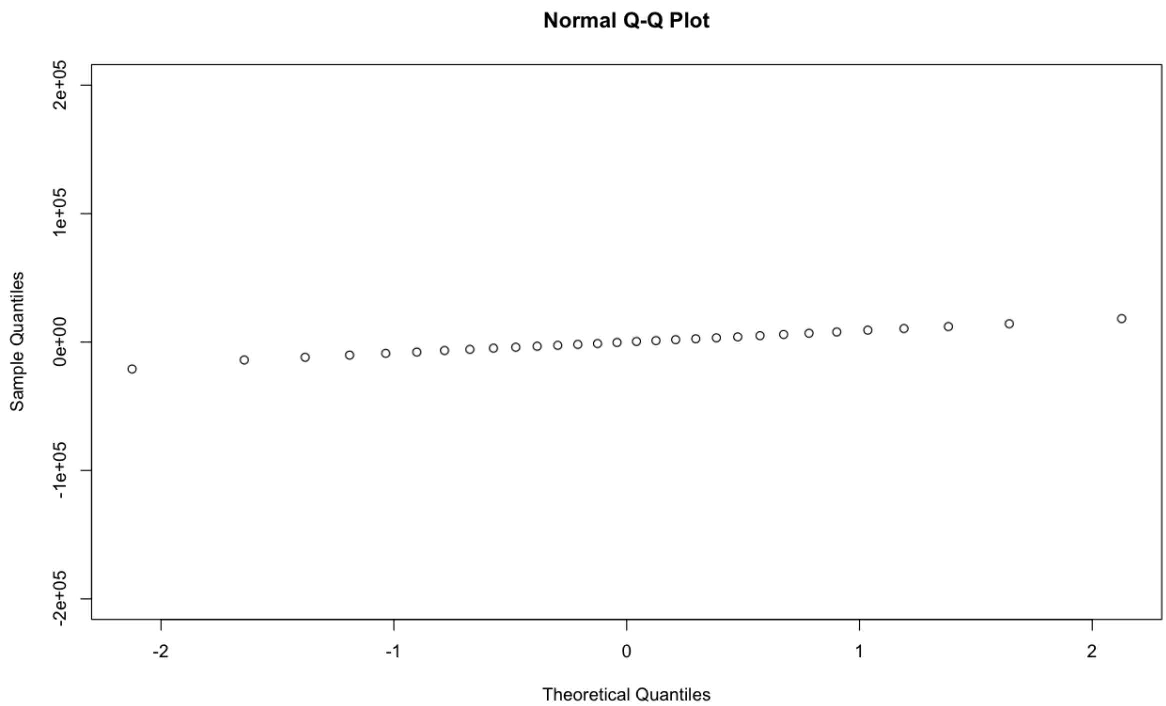

Figura 30. QQ plot de la predicción Banco Bogotá.

Se observa en la figura 30 el QQ plot del banco de Bogotá (normalidad de los residuales). Se puede observar que al comparar los cuantiles de la distribución normal estándar con los de los residuales se tiene un comportamiento lineal, lo cual es un buen indicio. Se realiza el test de Jarque- Bera y se obtiene un p-value = 0.22 donde confirmamos que los residuales se distribuyen normalmente.

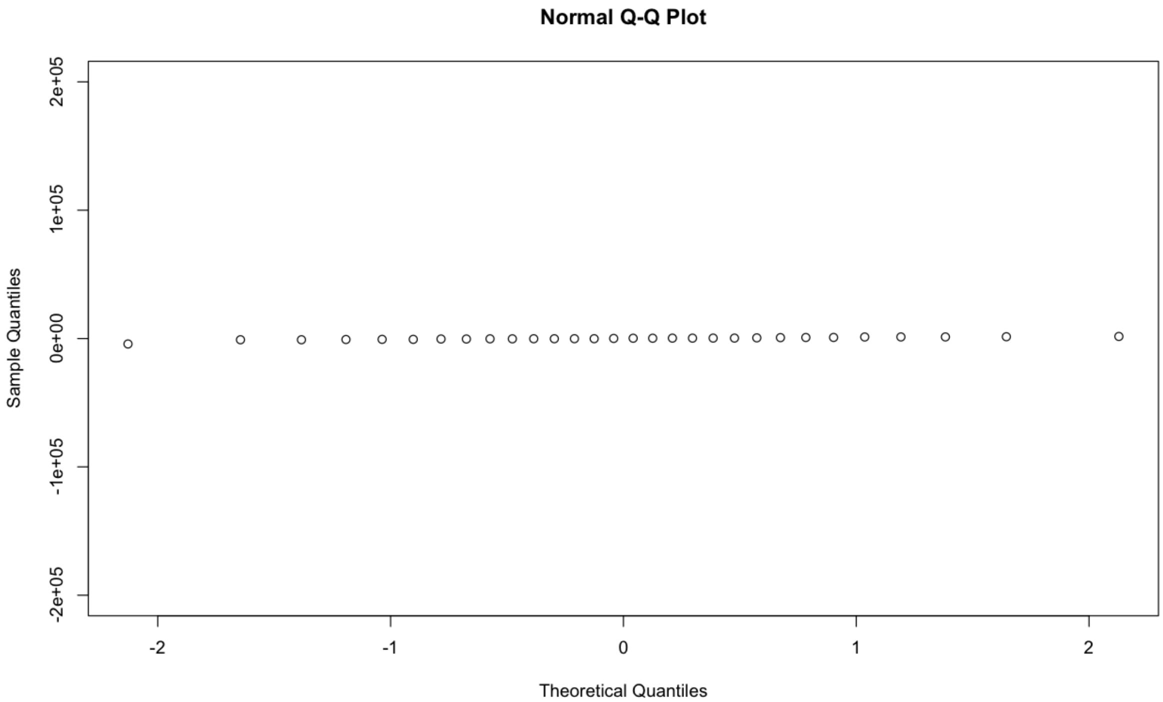

Figura 31. QQ plot de la predicción Banco de Occidente.

Con la figura 31 se evaluará la normalidad de los residuales en el banco de Occidente. Se puede observar en el anterior gráfico que al comparar los cuantiles de la distribución normal estándar con los de los residuales no hay un comportamiento lineal, mostrando que no hay buenos indicios de que se distribuyan normalmente.

10. CONCLUSIONES

- La metodología implementada se basa en un análisis temporal a partir de software estadístico R, que permite hacer predicciones de activos, en este caso los precios de las acciones de los tres bancos en consideración. Aunque es de resaltar que existen metodologías más robustas que permiten hacer pronósticos más certeros, estas superan con creces los fundamentos teóricos obtenidos durante los estudios. Sin embargo nuestro trabajo ejemplifica las técnicas básicas y permite calcular de forma explícita los intervalos de confianza en los que se podrían tener las estimaciones, obteniendo valores razonables.
- En la figura 29 - 30 - 31 se evalúa la normalidad de los residuales de las series de los bancos, Se puede observar que los Bancolombia y Bogotá, al comparar los cuantiles de la distribución normal estándar con los de los residuales se tiene un comportamiento lineal y esto es un buen indicio, el banco de occidente al comparar los cuantiles, no hay un comportamiento lineal, mostrando que no hay buenos indicios de que se distribuyan normalmente.
- El intervalo de máxima amplitud que permite la predicción de los activos en el futuro inmediatamente continuo a los datos de entrada va desde un mes hasta los dos meses. Con un mes se puede llegar a predecir hasta una semana de forma efectiva y para los dos meses de dos a tres semanas de forma efectiva.
- Se puede evidenciar en las tablas 4, 5 y 6 las predicciones de los activos de los bancos para la primera semana de Agosto, indicando un leve decaimiento en cada uno de estos activos y mostrando que no sería una buena opción invertir en dichos activos para la primera semana de agosto del año 2017.

- Para series no bien comportadas la predicción puede llegar a no ser tan precisa y se puede llegar a tener solo procesos autorregresivos, armas y si la serie es bien comportada en un periodo de tiempo se puede hasta obtener un arima.

  Tal como se puede evidenciar en las figuras 20, 24 y 28 del presente trabajo.

  Pues las mejores predicciones que se obtuvieron fueron de la forma ARIMA(p,0,0), tomando con base de la simulación el mes de julio de presente año (2017).

- Ver la efectividad de la predicción puede depender del pasado de la serie ya que al contar con la simulación de los valores pasados se supone que el futuro tendrá el mismo comportamiento.

- A lo largo del trabajo se pudo observar que los procesos ARIMA que mejor describían el comportamiento de las series de los bancos fueron los Procesos Autoregresivos (AR). Aunque no es sencillo en principio ver después de hacer las diferenciaciones respectivas para que el procesos sean estacionarios, qué modelo ARMA es el que mejor explica el comportamiento de los precios de las acciones, analizando las funciones fac y fpc, y con ayuda de los criterios del AIC y BIC que vienen incorporados en las librerías usadas, se pudo determinar que para este trabajo los procesos AR en su mayoría eran los que mejor explicaban estos comportamientos.

REFERENCIAS

[1], [8], [9], [10], [11], [13], [14] *Econometría*, quinta edición, 2009, *Damodar N. Gijarati, Dawn C. Porter,* página 740, página 749, página 753, página 754

[2], [3], [4] *Introduction to Probability and Stochastic Processes with Applications, Liliana Blanco Castaneda, Viswanathan Arunachalam Delvamuthu Dharmaraja*, 2012 página 342, página 473.

[5] *An Introduction to Stochastic Processes and Their Applications*, Cuadernos de Economía 48, 2008 *Petar Todorovic*, página 15, página 83.

SANCHEZ. P. (2008). Cambios estructurales en series de tiempo: una revisión del estado del arte. Revista de Ingenierías Universidad de Medellín, 6, pp. 115-140.

[6], [12] *Notas de Clase, Series de Tiempo con R, Norman Giraldo Gómez,* página 74, página 95

[7], *Análisis de Series de Tiempo para la Predicción de los Precios de la Energía en la Bolsa de Colombia, Sergio Botero Botero, Jovan Alfonso Cano Ca*no, página 176.

[15], *Package forecast, https://goo.gl/gtyA8V, [On-line] (2017), R. Hyndman*

[16], *Kincaid, D., Nathanksy, H. L. F., Alcaraz, C. T., Cheney, W., & Martínez, R. (1994). Análisis numérico: las matemáticas del cálculo científico.*